

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2003, 06:36:30 ; Search time 4701 Seconds

(without alignments)  
10883.372 Million cell updates/sec

Title: US-10-009-782-1

Perfect score: 1758

Sequence: 1 gaattccactgacgcgcga.....ccctgagctacgagaacct 1758

Scoring table: IDENTITY-NDC  
Gapop.10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

Score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1758	100.0	1758	6	BD011857	BD011857 Transform
2	1758	100.0	1758	6	E59056	E59056 Transformed
3	1758	100.0	1758	23	BD004254	BD004254 Transform
4	1739.4	98.9	1980	1	S80683	S80683 dan-D-amino
5	1453.4	82.7	1455	1	AFANADA	D45918 Alcaligenes
6	1088.6	61.9	1455	1	AF332548	AF332548 Alcaligen
7	693.6	39.5	1497	1	AFAG61	D45919 Alcaligenes
8	533.6	30.4	208050	1	AL646083	AL646083 Ralstonia
9	455.8	25.9	1467	1	AFAG6	D50061 Alcaligenes
10	306	17.4	110000	2	LMFCHR34_10	Continuation (11 o
11	171.8	9.8	22251	1	SC2K36	AL591857 Streptomy
12	116	6.6	82746	1	AF453501	AF453501 Actinosyn
13	107.4	6.1	28751	1	SC6A11	AL138978 Streptomy
14	99.2	5.6	14710	1	AE006004	AE006004 Caulobact
15	99.2	5.6	123580	6	AX211705	AX211705 Sequence
16	99.2	5.6	125401	6	AX211739	AF263912 Streptomy
17	97.2	5.5	113193	1	AF357202	AX211739 Sequence
18	94.2	5.4	5123	14	HEPVIE	AF357202 Streptomy
19	94.2	5.4	8438	14	SHILT	AF357202 Streptomy
20	94.2	5.3	34766	1	SCC121	M57505 Pseudorabie
21	94	5.3	4039	1	AF425991	AF425991 Streptomy
22	93.6	5.3	110000	2	LMFCHR32_06	AL137166 Streptomy
23	93.4	5.3	27541	6	AX211706	Continuation (7 of
24	91.4	5.2	37750	1	SC6D7	AX211706 Sequence
25	91.4	5.2	35033	1	SC130A	AL133213 Streptomy
26	91	5.2	35654	1	SC7H1	AL096811 Streptomy
27	91	5.1	15141	14	SHIPIOE	AL021411 Streptomy
28	89.8	5.1	33779	1	SCGD3	M34651 Pseudorabie
29	89.6	5.1	110000	2	LMFCHR36_31	AL096822 Streptomy
30	88.8	5.0	38404	1	SC2G5	Continuation (32 o
31	88.2	5.0	13848	1	AE005104	AL035478 Streptomy
32	87.8	5.0	1746	1	AF425992	AE005104 Halobacte
33	87.8	4.9	30657	1	SCD12A	AF425992 Streptomy
34	86.8	4.9	41906	1	SC5C7	AF357524 Streptomy
35	86.8	4.9	88421	6	AX417445	AL031515 Streptomy
36	85.6	4.9	5520	3	AC084329	AX417445 Sequence
37	85.6	4.9	49736	1	AF319998	AC084329 Leishman
38	85.2	4.8	25362	1	SCF56	AF319998 Stigmatel
39	84.6	4.8	109519	6	AX195929	AL133424 Streptomy
40	84.6	4.8	1839	6	AX196078	AX195929 Sequence
41	83.8	4.8	3578	1	AF005277	AX196078 Sequence
42	83.8	4.8	3619	1	SCVALSFP	AF005277 Cellulomo
43	83.8	4.8	24800	1	SC9B1	Y13070 S.coelicolo
44	83.8	4.8	110000	2	LMFCHR36_03	AL049727 Streptomy
45	83.8	4.8	110000	2	LMFCHR36_03	Continuation (4 of

#### ALIGNMENTS

RESULT 1  
BD011857  
LOCUS  
DEFINITION Transformant microorganism and process for producing  
ACCESSION BD011857 1758 bp DNA linear PART 02-AUG-2002  
VERSION BD011857.1 GI:22092046  
KEYWORDS D-aminocyclase.  
SOURCE unclassified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 1758)  
AUTHORS Takeuchi,K., Koide,Y., Hirose,Y., Moriyuchi,M. and Isobe,K.  
TITLE Transformant microorganism and process for producing D-aminocyclase  
JOURNAL Patent: WO 0078926-A 1 28-DEC-2000;

AMANO PHARMACEUTICAL CO LTD, KENICHI TAKEUCHI, YOSHINO KOIDE,  
YOSHIIHIKO HIROSE, MITSUAKI MORIGUCHI, KIMIYASU ISOBE  
OS Alcaligenes xylosoxydans subsp. xylosoxydans  
PN MO 0078926-A/1  
PD 28-DEC-2000  
PF 15-JUN-2000 WO 2000JP003932  
PI 17-JUN-1999 JP 99P 170555

PI KENICHI TAKEUCHI, YOSHINO KOIDE, YOSHIIHIKO HIROSE, PI  
MITSUAKI MORIGUCHI,  
PI KIMIYASU ISOBE  
PC C12N1/21, C12N15/52, C12N9/80  
CC  
FH Key Location/Qualifiers.

FEATURES  
source 1. 1758  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT. 281 a 662 c 568 g 247 t  
ORIGIN

Query Match 100.0% Score 1758; DB 6; Length 1758;  
Best Local Similarity 100.0%; Pred. No. 9e-183;  
Matches 1758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GAATTCACCTGATCGGGAGAGATTCATGTCCTCAATCCGATTCGAGCCCTTC 60  
|||||  
1 GAATTCACCTGATCGGGAGAGATTCATGTCCTCAATCCGATTCGAGCCCTTC 60  
61 GACCTGCTGTCGGGGGCGGACCTTCATCGACGGCAACACCCCGGGGCGGCGC 120  
|||||  
61 GACCTGCTGTCGGGGGCGGACCTTCATCGACGGCAACACCCCGGGGCGGCGC 120  
121 GACCTGGGCGTGGCGGCGGACCGCATCGCCCATCGGCGATCTGTGGAGCGCGCG 180  
|||||  
121 GACCTGGGCGTGGCGGCGGACCGCATCGCCCATCGGCGATCTGTGGAGCGCGCG 180  
181 CACACCGGGGTGAGCTGTCGGGGCTGTGGTGGCGCGCCGCTTCATGATCGACAC 240  
|||||  
181 CACACCGGGGTGAGCTGTCGGGGCTGTGGTGGCGCGCCGCTTCATGATCGACAC 240  
241 CACGACGACACTACCTCTCAGGCGTGCAGCATGACGCCCAAGATCTCGAGGGCG 300  
|||||  
241 CACGACGACACTACCTCTCAGGCGTGCAGCATGACGCCCAAGATCTCGAGGGCG 300  
301 ACCAGGTGTGACGGGCAATTTGGGGCATCAAGCTGGGGCGGCTGGCGCAACCCG 360  
|||||  
301 ACCAGGTGTGACGGGCAATTTGGGGCATCAAGCTGGGGCGGCTGGCGCAACCCG 360  
361 CCGGCGCCCTGAGCTCTGTGAGCAAGGCGGCTTACCGTTTGAGAGCGCTTGGCG 420  
|||||  
361 CCGGCGCCCTGAGCTCTGTGAGCAAGGCGGCTTACCGTTTGAGAGCGCTTGGCG 420  
421 TACCTGAGCGGCTTGGGCGGCAAGCGCGGCGGCGTCAAGCGCGCTGTATGGTGG 480  
|||||  
421 TACCTGAGCGGCTTGGGCGGCAAGCGCGGCGGCGTCAAGCGCGCTGTATGGTGG 480  
481 TCACGCTGCGGCGGCGGCTATGCGGACCTTGGAGCGCGCGCGCAAGCAAGAAATC 540  
|||||  
481 TCACGCTGCGGCGGCGGCTATGCGGACCTTGGAGCGCGCGCGCAAGCAAGAAATC 540  
541 GCGGCGATGCGGAGCTGCGGAGAGGCAATGGCAGGCGGCGGCTGGCGATTCGAGC 600  
|||||  
541 GCGGCGATGCGGAGCTGCGGAGAGGCAATGGCAGGCGGCGGCTGGCGATTCGAGC 600  
601 GCGGCGCTTACCGCGCGCGCGCGCGCGCGCGCACCAAGAGATCATGAGGTGGCG 660  
|||||  
601 GCGGCGCTTACCGCGCGCGCGCGCGCGCGCGCACCAAGAGATCATGAGGTGGCG 660  
661 CCGCTGAGGCGGATGGGCGGCTATCGCGCAACCGCATGCGGAGGAGGAGGACATC 720  
|||||  
661 CCGCTGAGGCGGATGGGCGGCTATCGCGCAACCGCATGCGGAGGAGGAGGACATC 720  
721 GTGGCGCGCTGAGGAACTTCGCGATGCGCGGAGACTGGAGTGGTGGATC 780

|||||  
721 GTGGCGCGCTGAGGAACTTCGCGATGCGCGGAGACTGGAGTGGTGGATC 780  
781 TCGCACCACCAAGGTGATGGGCGGACCCCAATTTGGCGGCTCGCGGAGAGCGTGGCG 840  
|||||  
781 TCGCACCACCAAGGTGATGGGCGGACCCCAATTTGGCGGCTCGCGGAGAGCGTGGCG 840  
841 ATGAGGCGGCGGATGGGCGGCGGAGGATGTCCTGAGCGGCTATCCCTACGTTGGCG 900  
|||||  
841 ATGAGGCGGCGGATGGGCGGCGGAGGATGTCCTGAGCGGCTATCCCTACGTTGGCG 900  
901 TCCACATGCTCAAGAGGAGCGGCGGCTGCTGGCGGAGGAGCAATCATCATCTGGTGC 960  
|||||  
901 TCCACATGCTCAAGAGGAGCGGCGGCTGCTGGCGGAGGAGCAATCATCATCTGGTGC 960  
961 AAGCCCTTCCCGCAATGAGCGGCGGCGGACCTGATGATGTCGCGCGGAGCGCGCAAA 1020  
|||||  
961 AAGCCCTTCCCGCAATGAGCGGCGGCGGACCTGATGATGTCGCGCGGAGCGCGCAAA 1020  
1021 TCCAGTACGACGTGTGTGCGGCGGAGCTGACCGCGGCGGCGGCTATCTACTATGATG 1080  
|||||  
1021 TCCAGTACGACGTGTGTGCGGCGGAGCTGACCGCGGCGGCGGCTATCTACTATGATG 1080  
1081 GAACCCGACGTCAGCGCATCTGCGGCTTGGCGCGGCGGAGCATGATCGGCTCGAGCG 1140  
|||||  
1081 GAACCCGACGTCAGCGCATCTGCGGCTTGGCGCGGCGGAGCATGATCGGCTCGAGCG 1140  
1141 CCGCAGCAGAGCGCGCGGCGGACCTTCGCGGCGGCGGCTTCCCGGCGGCGGCGG 1200  
|||||  
1141 CCGCAGCAGAGCGCGCGGCGGACCTTCGCGGCGGCGGCTTCCCGGCGGCGGCGG 1200  
1201 TATGCGGCGGACCTGCGGCTGTTCGCGGCTGAGAGCGGCTATGAGATGATCGGCG 1260  
|||||  
1201 TATGCGGCGGACCTGCGGCTGTTCGCGGCTGAGAGCGGCTATGAGATGATCGGCG 1260  
1261 ACCGCGGCGGCTTGGCGGCTGCGGCGGCGGCGGAGCTGAGGCGGCGGCTTTCGCG 1320  
|||||  
1261 ACCGCGGCGGCTTGGCGGCTGCGGCGGCGGCGGAGCTGAGGCGGCGGCTTTCGCG 1320  
1321 CTGGTGTGTTGACCGCGGCGGCGGAGTACCGCACTTCGCAACCTTACCGAG 1380  
|||||  
1321 CTGGTGTGTTGACCGCGGCGGCGGAGTACCGCACTTCGCAACCTTACCGAG 1380  
1381 CCGCGCGCGGCGGATCTGCGGCTGAGTCAAGCGGCGGCGGCTGAGGAGGAGGCG 1440  
|||||  
1381 CCGCGCGCGGCGGATCTGCGGCTGAGTCAAGCGGCGGCGGCTGAGGAGGAGGCG 1440  
1441 TTCACCGGCGGATGCGGCGGCGGCTGCGGAGCGGCGGCTGAGGAGGAGGAGG 1500  
|||||  
1441 TTCACCGGCGGATGCGGCGGCGGCTGCGGAGCGGCGGCTGAGGAGGAGGAGG 1500  
1501 CCGTTACATCGGCGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560  
|||||  
1501 CCGTTACATCGGCGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560  
1561 CCGTACATGCGGCGGCGGCTGCGGCTGAGTCAAGCGGCGGCGGCGGCGGCGG 1620  
|||||  
1561 CCGTACATGCGGCGGCGGCTGCGGCTGAGTCAAGCGGCGGCGGCGGCGGCGG 1620  
1621 GTGATGGGCGGCGGCTGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGG 1680  
|||||  
1621 GTGATGGGCGGCGGCTGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGG 1680  
1681 TCGAGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740  
|||||  
1681 TCGAGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740  
1741 CTGAGCTTACGAGGCTT 1758  
|||||  
1741 CTGAGCTTACGAGGCTT 1758

RESULT 2

ES9056  
 LOCUS 1758 bp DNA linear PAT 31-JAN-2002  
 DEFINITION Transformed microorganism and process for producing D-aminomylase.  
 ACCESSION E59056  
 VERSION E59056.1 GI:18629918  
 KEYWORDS JP 2001000185-A/1.  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE  
 1 (bases 1 to 1758)  
 Takeuchi, K., Koide, Y., Hirose, Y., Moriguchi, M. and Isobe, K.  
 Transformed microorganism and process for producing D-aminomylase  
 Patent: JP 2001000185-A 1 09-JAN-2001;  
 AMANO PHARMACEUT CO LTD  
 OS Alcaligenes xylosoxydans subsp. xylosoxydans  
 PN JP 2001000185-A/1  
 PD 09-JAN-2001  
 PF 17-JUN-1999 JP 1999170555  
 PR  
 PI KENICHI TAKEUCHI, YOSHINAO KOIDE, YOSHIHIKO HIROSE, PI  
 MITSUAKI MORIGUCHI,  
 PI KIMIYASU ISOBE  
 PC C12N15/09, C12N1/21, C12N9/80, (C12N15/09, C12R1:05), (C12N9/80,  
 PC C12R1:19),  
 PC C12N15/00, (C12N15/00, C12R1:05)  
 CC  
 CC  
 FH  
 FT  
 FT  
 FEATURES  
 source  
 Location/Qualifiers  
 1..1758  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 281 a 662 c 568 g 247 t  
 ORIGIN  
 Query Match 100.0%; Score 1758; DB 6; Length 1758;  
 Best Local Similarity 100.0%; Pred. No. 9e-183;  
 Matches 1758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 481 TCACGCTGCGCGCGCGCGCTATGCCGACTTGCAGCGCGCGCCACCGCAGAGAAATC 540  
 DB |||||||  
 DB 481 TCACGCTGCGCGCGCGCGCTATGCCGACTTGCAGCGCGCGCCACCGCAGAGAAATC 540  
 QY 541 GCGCCATGCGGACCTGGCGCGAGGAAGCATGGCGAGCGCGCCATCGCATTTTCGAC 600  
 DB |||||||  
 DB 541 GCGCCATGCGGACCTGGCGCGAGGAAGCATGGCGAGCGCGCCATCGCATTTTCGAC 600  
 QY 601 GCGCGCTTCTACCG 660  
 DB |||||||  
 DB 601 GCGCGCTTCTACCG 660  
 QY 661 CCGCTGAGCGCGCATGGCGCGCATCTACGCCACCCACATCGCGGAGAGAGAGAGAGATC 720  
 DB |||||||  
 DB 661 CCGCTGAGCGCGCATGGCGCGCATCTACGCCACCCACATCGCGGAGAGAGAGAGATC 720  
 QY 721 GTGGCGCGCGTGGAGAAACCTTCGCGCATGGCGCGCGCGCGCGCGCGCGCGCGCG 780  
 DB |||||||  
 DB 721 GTGGCGCGCGTGGAGAAACCTTCGCGCATGGCGCGCGCGCGCGCGCGCGCGCGCG 780  
 QY 781 TCGACACACAGATGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840  
 DB |||||||  
 DB 781 TCGACACACAGATGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840  
 QY 841 ATCGAGCGCGCGCATGGCG 900  
 DB |||||||  
 DB 841 ATCGAGCGCGCGCATGGCG 900  
 QY 901 TCCACATGCTACAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
 DB |||||||  
 DB 901 TCCACATGCTACAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
 QY 961 AAGCCCTTCCCGAAGCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020  
 DB |||||||  
 DB 961 AAGCCCTTCCCGAAGCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020  
 QY 1021 TCCAGTACAGAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080  
 DB |||||||  
 DB 1021 TCCAGTACAGAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080  
 QY 1081 GACCGGACGTGACAGCG 1140  
 DB |||||||  
 DB 1081 GACCGGACGTGACAGCG 1140  
 QY 1141 CCGCAGCAGAGCG 1200  
 DB |||||||  
 DB 1141 CCGCAGCAGAGCG 1200  
 QY 1201 TATGCGCGGACCTGGCG 1260  
 DB |||||||  
 DB 1201 TATGCGCGGACCTGGCG 1260  
 QY 1261 ACCGCGCGCGCTTGGCG 1320  
 DB |||||||  
 DB 1261 ACCGCGCGCGCTTGGCG 1320  
 QY 1321 CTGCTGTGTTGACAGCG 1380  
 DB |||||||  
 DB 1321 CTGCTGTGTTGACAGCG 1380  
 QY 1381 CCGCGCGCGCGCATATTCGCTGTACGTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440  
 DB |||||||  
 DB 1381 CCGCGCGCGCGCATATTCGCTGTACGTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440  
 QY 1441 TTCACGCGCGCATATTCG 1500  
 DB |||||||  
 DB 1441 TTCACGCGCGCATATTCG 1500  
 QY 1501 CCTTACATTCGCGGTGAAGCG 1560  
 DB |||||||  
 DB 1501 CCTTACATTCGCGGTGAAGCG 1560  
 QY 1561 CGCTACATTCGCGGTGCTCCCTCCGCTCGCATATACGCGCGCGCGCGCGCGCGCGCGCG 1620

RESULT 3

ID	BD004254	standard; DNA; UNC; 1758 BP.
XX	BD004254;	
SV	BD004254.1	
DT	08-FEB-2002 (Rel. 70, Created)	
Dt	08-FEB-2002 (Rel. 70, Last updated, Version 1)	
DE	Transformant microorganism and process for producing D-aminoacylase.	
KW	JP 03074329-T/1.	
XX	unidentified	
OS	unclassified	
OC	unclassified.	
XX	[1]	
RN	1-1758	
RA	Tateuchi K., Koide Y., Hirose Y., [Moriguchi M., Ito K.]	
RT	*Transformant microorganism and process for producing D-aminoacylase*;	
RL	Patent number JP03074329-T/1, 12-JAN-2001.	
XL	AMANO PHARMACEUTICAL CO LTD.	
XX		
CC	OS. Alcaligenes xylosoxydans subsp. xylosoxydans	
CC	CN JP 03074329-T/1	
CC	PD 12-JAN-2001	
CC	PF 15-JUN-2000 JP 2000003932	
CC	PR 17-JUN-1999 JP 99P 170555	
CC	PI KENICHI TAKEUCHI, YOSHINAO KOIDE, YOSHIHIKO HIROSE,	
CC	PI MITSUAKI MORIGUCHI,	
CC	PI KIMIYASU ISOBE	
CC	PC C12N1/21,C12N15/52,C12N9/80	
CC	CC CC	
CC	FH key Location/Qualifiers	
CC	FT source 1..1758	
CC	FT /db_xref="taxon:32644"	
CC	FT /organism="Alcaligenes xylosoxydans subsp."	
CC	FT xylosoxydans	
CC	XX key Location/Qualifiers	
CC	XX source 1..1758	
CC	XX /db_xref="taxon:32644"	
CC	XX /organism="unidentified"	
SO	Sequence 1758 BP; 281 A; 662 C; 568 G; 247 T; 0 other:	
Query Match	100.0%; Score 1758; DB 23; Length 1758;	
Best Local Similarity	100.0%; Pred. No. 9e-183;	
Matches 1758; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Oy	1 GAATTCACATTATCGCGAGGAGAAGATTTCCATGCCAATCCGATTCACGACCCTTC 60	
Dd	1 GAATTCACATTATCGCGAGGAGAAGATTTCCATGCCAATCCGATTCACGACCCTTC 60	
Oy	61 GACTTGCTCTCTCGGGGGGCACCTCATGACGGGAGCACACCCCGGGGGGGCGGCC. 120	

[illegible]



Db 1141 CCGACGACGAGCGCCGCGATCCGCGCTGTGGGCGACCTTCCCGGGGTGCTGGGGCAC 1200  
 QY 1201 TATGCGCGGACCTGGGCGCTGTCCCGTGAGAGGGGGTATGAGATGACCGGCGTG 1260  
 Db 1201 TATGCGCGGACCTGGGCGCTGTCCCGTGAGAGGGGGTATGAGATGACCGGCGTG 1260  
 QY 1261 ACCGCGCGCGCTTGGCGCTGGCGGGCGGGGACCTGACAGCGCGGATCTTCCGCGAC 1320  
 Db 1261 ACCGCGCGCGCTTGGCGCTGGCGGGCGGGGACCTGACAGCGCGGATCTTCCGCGAC 1320  
 QY 1321 CTGGGTGTGTGACACCGCGCGACGGTGGCCGATACCGCACTTGCAGACCTTACCGAG 1380  
 Db 1321 CTGGGTGTGTGACACCGCGCGACGGTGGCCGATACCGCACTTGCAGACCTTACCGAG 1380  
 QY 1381 CCGCGCGCGCGGACATTCCTGCTGACGTCACAGCGCGCGCTGTGGCAAGAGCGCGG 1440  
 Db 1381 CCGCGCGCGCGGACATTCCTGCTGACGTCACAGCGCGCGCTGTGGCAAGAGCGCGG 1440  
 QY 1441 TTTACCGGCGACGATGCGCGCGCGCTGTGCGACGACGCGCGCTGAGCGCGCGCGCAG 1500  
 Db 1441 TTTACCGGCGACGATGCGCGCGCGCTGTGCGACGACGCGCGCTGAGCGCGCGCGCAG 1500  
 QY 1501 CCCTTACATCCGCGCTGTAACGCGGGGCGCTGCGCCCTCCCAACCTTGACGCAAC 1560  
 Db 1501 CCCTTACATCCGCGCTGTAACGCGGGGCGCTGCGCCCTCCCAACCTTGACGCAAC 1560  
 QY 1561 CGCTACATGCGCGCTGCTCCGCTGCAATACGCGCGCGCGCTGCAATGCTGGGCAAGAA 1620  
 Db 1561 CGCTACATGCGCGCTGCTCCGCTGCAATACGCGCGCGCGCTGCAATGCTGGGCAAGAA 1620  
 QY 1621 GTGATGGGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680  
 Db 1621 GTGATGGGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680  
 QY 1681 TCGCAGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1740  
 Db 1681 TCGCAGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1740  
 QY 1741 CTGAGCTACGAGAGCTT 1758  
 Db 1741 CTGAGCTACGAGAGCTT 1758

## RESULT 4

S80683

LOCUS

1980 bp DNA linear BCT 13-AUG-1995  
 dan-D-aminocyclase [Alcaligenes xylosoxydans, ssp. xylosoxydans,

DEFINITION

S80683

ACCESSION

S80683.1

VERSION

S80683.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REMARK

FEATURES

source

gene

/gene="dan"

CDS  
 /note="D-aminocyclase, N-acyl-D-amino acid amidohydrolase"  
 201..1555  
 /gene="dan"  
 /note="3.5.1.14; Method: conceptual translation with  
 partial peptide sequencing; This sequence comes from Fig.  
 3; N-acyl-D-amino acid amidohydrolase"  
 /codon\_start=1  
 /transl\_table=1  
 /product="D-aminocyclase"  
 /protein\_id="AAB35881.1"  
 /db\_xref="GI:1488357"  
 /translation="MSQSDSPDLLLAGGTLIDGNSMTPGRADLGVRSGRNIAIGDL  
 SDAARHVRDVSGLVVAAGFTDSRTHDNDLRRDMPKISQVTVVYNGSISLA  
 PLANAPPAPDLIDEGSTFEFRADIDLRTPAVNAACWGHSTLAAAPDL  
 QRAVDEIEIAARLAEAMASGAIGSTGAFPPARATTEIEICRLSAHGIIY  
 ATHMDDSEHVALETFRIIGRELDPVIVSHKVGOPNFGRETLLEAMAR  
 ODVELDAYPYVAGTMDKODVLLAGRTIITWCKPELSELDDEVAERGSXYV  
 VPELOPAGAIYFMDEPVOIRILAFGPTMIGSDGLPHDERHPRLMCTFPRIYGVYR  
 DLGLEPETAVMKMGILTAARFGLAGSGLOAGYFADLVYFDPATVADTATFEHPTER  
 AAGHSYVYVNAFPWQEDATFGQAGRYLARTAA"  
 BASE COUNT 298 a 758 c 642 g 282 t  
 ORIGIN  
 Query Match 98.9%; Score 1739.4; DB 1; Length 1980;  
 Best Local Similarity 99.4%; Pred. No. 9.1e-181;  
 Matches 1746; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 GAATTCACCTGATGCGGAGAGAGATTGATGTCCTCAATCGATTCCAGCCCTTC 60  
 Db 168 GATGCCCACTGATGATGCCCGCGCGGAGATTTCATTCCTCAATCGATTCCAGCCCTTC 227  
 QY 61 GACGTGCTGCTGCGGGGGGACCCCTCATCGAGCGGACCAACACCCCGGGCGGGCCCC 120  
 Db 228 GACCTGCTGCTGCGGGGGGACCCCTCATCGAGCGGACCAACACCCCGGGCGGGCCCC 287  
 QY 121 GACCTGCGGCTGCGGGGACCGCATCGCCGCGCATGCTGTCGAGCGCGCGCGCG 180  
 Db 288 GACCTGCGGCTGCGGGGACCGCATCGCCGCGCATGCTGTCGAGCGCGCGCGCGCG 347  
 QY 181 CACACCCGGGTGACGTGTGCGGGCTGTGTCGCGCCCGGCTTCATGACGTGCACAC 240  
 Db 348 CACACCCGGGTGACGTGTGCGGGCTGTGTCGCGCCCGGCTTCATGACGTGCACAC 407  
 QY 241 CACGACGACAACTACTCTCTAGCGGTGCGGACATGAGCGCCCAAGATTCGCAAGGCGTC 300  
 Db 408 CACGACGACAACTACTCTCTAGCGGTGCGGACATGAGCGCCCAAGATTCGCAAGGCGTC 467  
 QY 301 ACCACGCTGTGTCACGGGCAATTGCGGCTACGCTGCGCGCGCGCGCGCGCGCGCG 360  
 Db 468 ACCACGCTGTGTCACGGGCAATTGCGGCTACGCTGCGCGCGCGCGCGCGCGCGCGCG 527  
 QY 361 CCGCGCCCGCTGACCTCTGTGACGAGAGCGGCTTACCGTTTGACGCGCTTGCAGC 420  
 Db 528 CCGCGCCCGCTGACCTCTGTGACGAGAGCGGCTTACCGTTTGACGCGCTTGCAGC 587  
 QY 421 TACCTGAGCGGCTTTCGGGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
 Db 588 TACCTGAGCGGCTTTCGGGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 647  
 QY 481 TCAACGCTGCGCGCGCGCGCTGATCGCGGCTTTCGAGCGCGCGCGCGCGCGCGCGCG 540  
 Db 648 TCAACGCTGCGCGCGCGCGCTGATCGCGGCTTTCGAGCGCGCGCGCGCGCGCGCGCG 707  
 QY 541 CCGCGCATGCGGCG 600  
 Db 708 CCGCGCATGCGGCG 767  
 QY 601 GCGCGCTTCTACCG 660  
 Db 768 GCGCGCTTCTACCG 827  
 QY 661 CCGCTGACG 720

Db 828 CCGTGAACGGCGATGCGGATCTACGCCACCCATGCCGACGAAGCGGACACATC 887  
 Qy 721 GTGGCCGGCTGAGAGAAACCTTCGCAATCGCCGCGAGCTGGAAGTCCGGTGGTATC 780  
 Db 888 GTGGCCGGCTGAGAGAAACCTTCGCAATCGCCGCGAGCTGGAAGTCCGGTGGTATC 947  
 Qy 761 TCGACCAACAAAGTATGATGAGCCCAATTTCCGCGCTCGCCGAGACCTGCGCTG 840  
 Db 948 TCGACCAACAAAGTATGATGAGCCCAATTTCCGCGCTCGCCGAGACCTGCGCTG 1007  
 Qy 841 ATGAGGCGCGCATGAGCGCGGAGAGCTTCGCTGAGAGGGTATCCCTAGTGGCGGC 900  
 Db 1008 ATGAGGCGCGCATGAGCGCGGAGAGCTTCGCTGAGAGGGTATCCCTAGTGGCGGC 1067  
 Qy 901 TCCACCATGCTCAACAGACGACCGCTGCTGCTGAGCGGAGCACCATCATCTGATG 960  
 Db 1068 TCCACCATGCTCAACAGACGACCGCTGCTGCTGAGCGGAGCACCATCATCTGATG 1127  
 Qy 961 AAGCCCTTCCCGAAGTATGAGCGGCGGAGCTGATGAATGCGCGCGAGACCGGCA 1020  
 Db 1128 AAGCCCTTCCCGAAGTATGAGCGGCGGAGCTGATGAATGCGCGCGAGACCGGCA 1187  
 Qy 1021 TCCAGTATGAGCTGATGAGCGGCGGAGCTGATGAATGCGCGCGAGACCGGCA 1080  
 Db 1188 TCCAGTATGAGCTGATGAGCGGCGGAGCTGATGAATGCGCGCGAGACCGGCA 1247  
 Qy 1081 GAACCCGAGCTGACAGCGGATCTGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1140  
 Db 1248 GAACCCGAGCTGACAGCGGATCTGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1307  
 Qy 1141 CCGCAGAGAGAGCGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1200  
 Db 1308 CCGCAGAGAGAGCGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1367  
 Qy 1201 TATGCGCGGAGCTGAGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1260  
 Db 1368 TATGCGCGGAGCTGAGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1427  
 Qy 1261 ACCGCGCGGAGCTGAGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1320  
 Db 1428 ACCGCGCGGAGCTGAGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1487  
 Qy 1321 CAGGCGGAGCTGAGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1380  
 Db 1488 CAGGCGGAGCTGAGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1547  
 Qy 1381 CCGCGCGGAGCTGAGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1440  
 Db 1548 CCGCGCGGAGCTGAGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1607  
 Qy 1441 TTCACCGGAGCTGAGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1500  
 Db 1608 TTCACCGGAGCTGAGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1667  
 Qy 1501 CCCTTACATCGGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1560  
 Db 1668 CCCTTACATCGGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1727  
 Qy 1561 CGCTACATGCGGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1620  
 Db 1728 CGCTACATGCGGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1787  
 Qy 1621 GTGATGAGCGGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1680  
 Db 1788 GTGATGAGCGGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1847  
 Qy 1681 TCGCAGAGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1740  
 Db 1848 TCGCAGAGCGGAGCTGAGCGGAGCTGAGCGGAGCAGATGATGCTCGAGCGCTG 1907  
 Qy 1741 CTGAGCTAGAGAGCT 1757  
 Db 1908 CTGAGCTAGAGAGATT 1924

RESULT 5  
 AFANADA  
 LOCUS  
 DEFINITION  
 accession  
 version  
 keywords  
 source  
 ORGANISM  
 Achromobacter xylosoxidans  
 Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 Achromobacter.  
 1 (sites)  
 Makayama, M., Ashika, T., Miyamoto, Y., Yoshikawa, T., Sonoda, Y.,  
 Sakai, K. and Moriiguchi, M.  
 Primary structure of N-acetyl-D-glutamate amidohydrolase from  
 Alcaligenes xylosoxidans subsp. xylosoxidans A-6  
 J. Biochem. 118 (1), 204-209 (1995)  
 96015170  
 2 (sites)  
 Makayama, M., Katsuno, Y., Hayashi, S., Miyamoto, Y., Sakai, K. and  
 Moriiguchi, M.  
 Cloning and sequencing of a gene encoding D-aminoacylase from  
 Alcaligenes xylosoxidans subsp. xylosoxidans A-6 and expression of  
 the gene in Escherichia coli  
 Biosci. Biotechnol. Biochem. 59 (11), 2115-2119 (1995)  
 96100942  
 3 (bases 1 to 1455)  
 Makayama, M.  
 Unpublished  
 4 (bases 1 to 1455)  
 Makayama, M.  
 Direct Submission  
 Submitted (13-FEB-1995) Mamoru Makayama, Oita University, Applied  
 Chemistry, Dannoanaru, Oita, Oita 870-11, Japan  
 (tel:0975-69-3311(ex.746), fax:0975-69-7957)  
 location/Qualifiers  
 1. 1455  
 /organism="Achromobacter xylosoxidans"  
 /isolate="A-6"  
 /sub\_species="xylosoxidans"  
 /db\_xref="taxon:85698"  
 1. 1455  
 /codon\_start=1  
 /transl\_table=1  
 /product="N-acetyl-D-amino acid amidohydrolase"  
 /protein\_id="BA08349.1"  
 /db\_xref="GI:1881397"  
 /translation="MSQSDSQPEDLLAGTLLIDGNTPRRARDLGVGRDRIAIQDL  
 SDAARHVRVSGIVAFGEIDSHDNDLRLRRDPTKISQGVTVVYVNGGISLA  
 PLAAHPAPDLIDEGSYRFEERADYDLRLRTPAIVNAACVNGHSTLRRAVMDL  
 ORAAHDEIIMARLAEAMASGAGIGTAAPAPRAARTEELITVCGRLSHAGIG  
 ATRHRDEHIVALEETFRIGREDLPVYVISHKVKGNRGRSRETLPLITAMAR  
 ODVSLDAIVPVAGSTMLKQDVLVLAGFTLLTWCKPFPELSGRUDDEVAAERKSKTDV  
 VPELOPAGAIYFMNDEPVDORILAFGTITMGSGPLPDERPHRLMVTFFRVIGHYAR  
 DLGFLPELTAVMKWTGLTAARFGIAGRGQIAGYFADLVVDPATVADTATFEPHPTER  
 AAGISHYVYVNAAPYMOEAFQGHAGVLAFTAA"  
 BASE COUNT 222 a 552 c 477 g 204 t  
 ORIGIN  
 Query Match 82.7%; Score 1453.4; DB 1; Length 1455;  
 Best Local Similarity 99.9%; Pred. No. 1.3e-149;  
 Matches 1454; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 34 ATGTCCCAATCGGATTCAGCGCTTCACTGCTGCTCGGCGGAGCACCCTCATGAC 93  
 Db 1 ATGTCCCAATCGGATTCAGCGCTTCACTGCTGCTCGGCGGAGCACCCTCATGAC 60  
 Qy 94 GGCAGCAACCCCGGCGGCGCGGAGCTGAGCGGAGCAGCGGAGCAGCGGCGCC 153

Dp	61	GGAGAGCAACACCCCGGGGGGGGGGGGGGCGGACACTGTGGGCGTGGGGGAGACGCATCCCGCC	120
Oy	154	ATCGGCATCTGTGAGACGCCCGCGCACACCUGGGTGCAGCTGTGGGCTGTGTGTC	213
Dp	121	ATCGGCATCTGTGAGACGCCCGCGCACACCUGGGTGCAGCTGTGGGCTGTGTGTC	180
Oy	214	GGCGCGGGCTTCACTGCATGGCATGCCAACGACGAACTACTCTGTCAAGGTGTGCAC	273
Dp	181	GGCGCGGGCTTCACTGCATGGCATGCCAACGACGAACTACTCTGTCAAGGTGTGCAC	240
Oy	274	ATGACGCCCAAGATCTGTGCAGAGGCGTACACAGCGTGTGCAGGCAATTTGGCGATACG	333
Dp	241	ATGACGCCCAAGATCTGTGCAGAGGCGTACACAGCGTGTGCAGGCAATTTGGCGATACG	300
Oy	334	CTGGCGCGCGTGGCGCGCACGCCAACCCGCCGCCCCCTGTGAGCTGTGCAGAGAGCGGC	393
Dp	301	CTGGCGCGCGTGGCGCGCACGCCAACCCGCCGCCCCCTGTGAGCTGTGCAGAGAGCGGC	360
Oy	394	TCCTTACGTTTCGAGCGCCTTGGCGGACACTGTGAGACGCTTGCGGGCGCACGCCGGCGC	453
Dp	361	TCCTTACGTTTCGAGCGCCTTGGCGGACACTGTGAGACGCTTGCGGGCGCACGCCGGCGC	420
Oy	454	GTCACACCCCGCTTATATGTGTGGGCGCAATTAACGCTCGCGCGCGCGTCAATGCGGACTTG	513
Dp	421	GTCACACCCCGCTTCAATGTGTGGGCGCAATTAACGCTCGCGCGCGCGTCAATGCGGACTTG	480
Oy	514	CAGCGCGCGCGCACCGACGAGAAATTCGGCGCCATCGGAGACTGTGACGAGAGAACCATG	573
Dp	481	CAGCGCGCGCGCACCGACGAGAAATTCGGCGCCATCGGAGACTGTGAGAGAACCATG	540
Oy	574	GGCAGCGCGCGCAATCGGCATTTTGACCCGGGCGCTTACC CGCGCGCGCGCGCGCAC	633
Dp	541	GGCAGCGCGCGCAATCGGCATTTTGACCCGGGCGCTTACC CGCGCGCGCGCGCGCAC	600
Oy	634	ACCGAAGATCATCGAGGTGTCCGGCGCTGAGCGCGCATGGCGGCAATTCAGCCAC	693
Dp	601	ACCGAAGATCATCGAGGTGTCCGGCGCTGAGCGCGCATGGCGGCAATTCAGCCAC	660
Oy	694	CACATCGCGCAGAGAGGCGAGCACATCTGGCGCGCTGGAGAAACCTTCCGCAATCGC	753
Dp	661	CACATCGCGCAGAGAGGCGAGCACATCTGGCGCGCTGGAGAAACCTTCCGCAATCGC	720
Oy	754	CGCGAGCTGAGACGTGCCGGGTGTGATCTGCACACAAAGTCAATGGGCAAGCCCAATTC	813
Dp	721	CGCGAGCTGAGACGTGCCGGGTGTGATCTGCACACAAAGTCAATGGGCAAGCCCAATTC	780
Oy	814	GGCGCTGCGCGCAGAGCGTGGCGGATGAGAGCGCGCATGGCGCGCGCAGACCTCG	873
Dp	781	GGCGCTGCGCGCAGAGCGTGGCGGATGAGAGCGCGCATGGCGCGCGCAGACCTCG	840
Oy	874	CTGAGACGATCTCTACGTGTGAGCGGCTCCACCATGCTCAACGAGAGCGCTGTGTG	933
Dp	841	CTGAGACGATCTCTCTACGTGTGAGCGGCTCCACCATGCTCAACGAGAGCGCTGTGTG	900
Oy	934	GCCGAGCGCACATCATCACTGTGTGAAAGCCCTTCCCGAACTAGCGGGCGCACCTG	993
Dp	901	GCCGAGCGCACATCATCACTGTGTGAAAGCCCTTCCCGAACTAGCGGGCGCACCTG	960
Oy	994	GATGAGTGGGGGGCGGCGCGCAATCAAGTACGAGAGTGTCCGAGCTGTGACCG	1053
Dp	961	GATGAGTGGGGGGCGGCGCGCAATCAAGTACGAGAGTGTCCGAGCTGTGACCG	1020
Oy	1054	GCGCGCGCATCTTACTTCATGATGAGAGAACCGGAGCTGACAGCATCTGTGGCTGGC	1113
Dp	1021	GCGCGCGCATCTTACTTCATGATGAGAGAACCGGAGCTGACAGCATCTGTGGCTGGC	1080
Oy	1114	CCGACCATGATCGGCTCCGACGGCTGTGCGCAGCAGACGCGCCGCAATCCGCGCTGTGG	1173
Dp	1081	CCGACCATGATCGGCTCCGACGGCTGTGCGCAGCAGACGCGCCGCAATCCGCGCTGTGG	1140
Oy	1174	GGCAGCTTCCCGGGGGGTGAGGGGCAATATGCGCGGAGACTGTGGGCTGTGCCGTGGAG	1233
Dp	1141	GGCAGCTTCCCGGGGGGTGAGGGGCAATATGCGCGGAGACTGTGGGCTGTGCCGTGGAG	1200

QY	1234	ACGGGGGGTATGTAAGATGACCGGGCCCTGACACCGCGGGCGCTTGGCGCTGCGCGGGCGGG	1293
QY	1234	ACGGGGGGTATGTAAGATGACCGGGCCCTGACACCGCGGGCGCTTGGCGCTGCGCGGGCGGG	1293
Db	1201	ACGGGGGGTATGTAAGATGACCGGGCCCTGACACCGCGGGCGCTTGGCGCTGCGCGGGCGGG	1260
QY	1294	CAGCTGCAGAGCGGGGTACTTTCGCGCACCTGTGTGTGTGTGACACCGGACAGCGGTGGCCGAT	1353
Db	1261	CAGCTGCAGAGCGGGGTACTTTCGCGCACCTGTGTGTGTGTGACACCGGACAGCGGTGGCCGAT	1320
QY	1354	ACCGCCACCTTGAACACCCCTACCGAGCGCGCGCCGCGGCATTCATTCGCTGACGTCAAC	1413
Db	1321	ACCGCCACCTTGAACACCCCTACCGAGCGCGCGCCGCGGCATTCATTCGCTGACGTCAAC	1380
QY	1414	GGCGGGCGCGGTGTGGCAAGAGACGAGCGGTTCACCGGGCCAGCATGCCGGCGCGCTGTCCCA	1473
Db	1381	GGCGGGCGCGGTGTGGCAAGAGACGAGCGGTTCACCGGGCCAGCATGCCGGCGCGCTGTCCCA	1440
QY	1474	CGCAGCGCGCGCTGCA 1488	
Db	1441	CGCAGCGCGCGCTGCA 1455	
RESULT 6	AF332548	1455 bp DNA linear BCT 04-MAR-2001	
LOCUS	AF332548		
DEFINITION	Alcaligenes faecalis D-aminooxylase gene, complete cds.		
ACCESSION	AF332548		
VERSION	AF332548.1	GI:13194735	
KEYWORDS	Alcaligenes faecalis.		
SOURCE	Alcaligenes faecalis.		
ORGANISM	Bacteria: Proteobacteria; beta subdivision; Alcaligenaceae;		
REFERENCE	1 (bases 1 to 1455)		
AUTHORS	Hsu, C.S., Lai, W.L., Chang, W.W., Yang, Y.B. and Tsai, Y.C.		
TITLE	Gene cloning, nucleotide sequencing of Alcaligenes faecalis DAI		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1455)		
AUTHORS	Hsu, C.S., Lai, W.L., Chang, W.W., Yang, Y.B. and Tsai, Y.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-Dec-2000) Institute of Biochemistry, National		
FEATURES	Yang-Ming University, 115, Li-Nong St. Sec. 2, Shih-Pai, Taipei,		
Source	Taiwan 11221, R.O. China		
Location/Qualifiers	1. 1455		
/organism="Alcaligenes faecalis"			
/db_xref="taxon:511"			
1. 1455			
/EC_number="3.5.1.81"			
/note="Daaf"			
/codon_start=1			
/transl_table=1			
/product="D-aminooxylase"			
/protein_id="AAK15530.1"			
/db_xref="GI:13194736"			
/translation="MSQPDATPPPYILSGGVYDGTNAPGLADVYRGDRIAVGDID			
SASSAKRIDVAGVVSSEGTIDSHDNDYTLKRRDTPKISQVTVTVYVNCISLILA			
PLAANPAPAPLDLDEGGSEFFARFSEYLELRLAPPAVNAACWGHSTLEAAVMPDID			
RRRTADEIQMQALADALASGAIGISTGAFYPAHAASFEELIEVCRPIITGGVYV			
ATNARDEGEHIVQALEEFTFRIGRELDVYVSHHKVMGKLNGRSKERTALIEAAMS			
QDVSLAPYVYAGSTMLKODRVLLAGRTLITWCKDPYELSGRDLIELTAERGSKYDV			
VPEIQPAGAIYFMNDEPDYVITLAFGRPHMISDCLPHDERHRLMKGTFFPVLYHSR			
DGLIPELTAVKMTGILTAKEFGLAERKQDPGYVADLVFDPATVADSATSEHPTEER			
AAGHSYVYVNAAMVEDQSTFGAGRVLNAGA"			
BASE COUNT	247 a 537 c 446 g 225 t		
ORIGIN			
Query Match	61.9%	Score 1088.6;	DB 1; Length 1455;
Best Local Similarity	84.38;	Prod. No. 6.6e-110;	
Matches 1226; Conservative	0;	Mismatches 229;	Indels 0; Gaps 0;
QY	34	AGTCCCAATCCGATCCAGTCCAGCGCTTGCAGCGTCTGCTGCGGGCGGACCGCTCATCGAC	93

D	b	1	ATGTCCACGCCGACGCCACCCGCTTTCGACTACATCTTTCCGGGGGACCGTATCGAC	60
O	y	94	GGCAGCAACACCCCGGGGGGGCGCCGACCTGTGGCGGGGAGACCGATGGCCGC	153
D	b	61	GGCAGCAACCGCCGACGCCGGGCTCGCCGACGTGGGCGTGGGGGATGGATTGGCCGC	120
O	y	154	ATCGGCGATCTGTGGAGCGCCGCGGCGCACCCCGGGGTGAGTGTGGGCTGTGGTC	213
D	b	121	GTGCGGACCTTAGGCGCCAGACAGCGCCCGCGCGCATCGAGCTGGGGGAAAGGTGTG	180
O	y	214	GCGCCGGGCTTACGACTTCGCACACCCAGCAGCAACTACCTACTGAGCGTGGAC	273
D	b	181	TTCTCGGGGCTTCTACCTACCTGCGACACGCACTGACCAACTCTGTGAAACACCGGAC	240
O	y	274	ATGACGCGCCAAAGTCTCGAGAGGGCGTACACAGGGGTGACGGGCAATTTGGGACTAGC	333
D	b	241	ATGACGCGCCAAAGTCTCGAGAGGGCGTACACAGGGGTGACGGGCAATTTGGGACTAGT	300
O	y	334	CTGGCGCGCTGTGGCGACGCCCAACCCGCGCCCGCTGTGGACTTGTGAGAGAGCGGC	393
D	b	301	CTGGCGCGCGCTGTGGCGACGCCCAATCGCGCGCGGCGTGTCTGTGAGAGAGCGGC	360
O	y	394	TCCTTACCGTTTGGAGCGCTTTCGCCGACTACTGAGACGGCTTGGGGCCAGCGCGGCGC	453
D	b	361	TCGTTCCGCTTTCACAGCTTCTCGGACTACTGAGAGCCCTTGGCGCGCGCGCTGGCC	420
O	y	454	GTCAGCGCGCGCTGTATGGTGGGGCACTTCAAGCGTGGCGCGCGCGCTATCGGACTTG	513
D	b	421	GTCAGCGCGCGCTGTATGGTGGGGCACTTCAAGCGTGGCGCGCGCGCTATCGGACTTG	480
O	y	514	CAGCGCGCGCGCACCGACGAGGAAATCGCGGCGATGCGGGAGACTTGGCGAGAGACCATG	573
D	b	481	CGGCGGCAACCCAGCGCGCGATGAAATCCAGGCGATGCAAGGGCGTGGCGAGATGGCGTG	540
O	y	574	GCCAGCGGGCGCATCGGCATTTGACCGGCGCTTCTACCGCGCGCGCGCGCGCGCAC	633
D	b	541	GCCAGTGGCGCATCGGCATTTTCACAGGCGCGCTTCTACCGCGCGCGCGCGCATGCGAC	600
O	y	634	ACCGAAGATGATCATGAGGTGGCGGCGGAGCGGAGCGGATGTACGGCAC	693
D	b	601	ACCGAAGATGATCATGAGGTGGCGGCGGAGCGGAGCGGATGTACGGCAC	660
O	y	694	CACATGCGCGACAGAGGCGAGCACATCGTGGCGGCGTGGAGGAACCTTCCGATCGGC	753
D	b	661	CATATGGCGGATGAGGGCGGAATATCGTCCAGGGCGTGGAGGAGAACCTTCCGATCGGC	720
O	y	754	CGCGAGCTGAGACTGCCGGTGGTATCTGCAACCAAGAGTATGGGAGCCCAATTTTC	813
D	b	721	CGCGAATGAGCACTGGCGGCTGTATCTGCAACCAAGAGTATGGGAGCCCAATTTTC	780
O	y	814	GGCGCGCGCGGAGAGCGGCGCGCGTGAATCGAGGGCGCGCATGGCGCGGAGAGAGTCTCG	873
D	b	781	GGGCGCTTCAGAGAAACGCTGGCTTGTATCGAGGGCGCGCGCATGGCGCGGAGAGTCTCG	840
O	y	874	CTGAGCGCTATCTCTACAGTGGCGGCTCCACATATGCTACAGAGACGGGTCTGCTG	933
D	b	841	CTGAGCGCTATCTCTACAGTGGCGGCTCCACATATGCTACAGAGACGGGTCTGCTG	900
O	y	934	CGCGGAGCGCACTCATACCTGTGTGCAAGCCCTTCCCGAACTGAGCGGGCGCGACTG	993
D	b	901	GCGCGCGCGCACTGTATCTGTGTGCAAGCCCTTCCCGAACTGAGCGGAGCGGACTG	960
O	y	994	GATGATATGCGCGGCGGAGCGCGCGCAATATCAAGTACAGAGTGGTCCCGCATGCAAGCG	1053
D	b	961	GATGATATGCGCGGCGGAGCGCGCGCAATATCAAGTACAGAGTGGTCCCGCATGCAAGCG	1020
O	y	1054	GCGCGCGCATCTACTTCAATGATGAGAGCAACCGAGTGCAGACCGCATCTGTGGCTGGC	1113
D	b	1021	GCGCGCGCATCTACTTCAATGATGAGAGCAACCGAGTGCAGACCGCATCTGTGGCTGGC	1080
O	y	1114	CGGACATGATGGCTTCGAGCGGCTTGGCGGACGAGACGCGCGCATTCGGGCTGTGG	1173

Db	1081	CCACCATGATGACGGGTCCGATGAGCCTGCCCAACGACGAGCGCCCCACCCACTCGGCTGTGG	1140
QY	1174	GGACACTTCCCGCGGGTGTCTGGGCACTATGCGCGACACTGGGCTGTTCCTCCGCTGGAG	1233
Db	1141	GGAACTTTCCTCCCGGTCTCTGGGCCACTACTCTCGCGACCTGGGCTCTTTCCTCAATGGAG	1200
QY	1234	ACGGCGGTATGGAAAGTAGACCGGCCTGACCGCCGCGCGCTTTCGGCTTGGCCGCGGCGGG	1293
Db	1201	ACGGCGGTCTGAAATATGACCGGCTTGACCGCCGCAATATTTCGGCTGTGCGGAACGGGGC	1260
QY	1294	CAGCTGCAGCGCCGGGTACTTTCGCGGACCTGGTGGTGTTCGACCCGGGCAACGGTGGCGGAT	1353
Db	1261	CAGGTGCAGCCCGGTCTACTAGCCGACCTGGTGGTGTTCGACCCGGGCAACGGTGGCGGAT	1320
QY	1354	ACCGGCACACTTCGAACACACTTACGAGGCGCGCGGCGGCAATCCATTGGTGTAGGTCAAC	1413
Db	1321	TGCGGCACGTTTCGACATATCCCAACGAGGCGCGCGGCGGCAATCCATTGGTGTAGGTCAAC	1380
QY	1414	GGCGGCGCGGCTGTGCGCAAGACAGAGCGTTTCAACCGGCGACGATGCGCGGCGGCTGTGCGCA	1473
Db	1381	GGCGGCGCGGCTGTGCGGAAGACACAGTCTTCACTGTGCCACGATGCGCGGCGGCTGTCAAC	1440
QY	1474	CGCACGGCGCGCTGA	1488
Db	1441	CGGGCGGAGGCTGA	1455

RESULT 7	
LOCUS	AFAA61
DEFINITION	1497 bp DNA linear BCT 10-FEB-1999
ACCESSION	AF045919
VERSION	DA5919.1 GI:1136626
KEYWORDS	N-acyl-D-amino acid amidohydrolase.
SOURCE	Alcaligenes xylosoxydans (sub-species:xylosoxydans, isolate:A-6) DNA.
ORGANISM	Achromobacter xylosoxydans
REFERENCE	Bacteria: Proteobacteria; beta subdivision; Alcaligenaceae; Achromobacter.
AUTHORS	1 (sites) Wakayama,M., Ashika,T., Miyamoto,Y., Yoshikawa,T., Sonoda,Y., Sakai,K. and Moriguchi,M.
TITLE	Primary structure of N-acyl-D-glutamate amidohydrolase from Alcaligenes xylosoxydans subsp. xylosoxydans A-6
JOURNAL	J. Biochem. 118 (1), 204-209 (1995)
MEDLINE	96015170
PUBMED	8537313
REFERENCE	2 (sites) Wakayama,M., Watanabe,E., Takenaka,Y., Miyamoto,Y., Tau,Y., Sakai,K. and Moriguchi,M.
AUTHORS	Cloning, expression, and nucleotide sequence of the N-acyl-D-aspartate amidohydrolase gene from Alcaligenes xylosoxydans subsp. xylosoxydans A-6
TITLE	J. Ferment. Bioeng. 80, 311-317 (1995)
JOURNAL	3 (bases 1 to 1497)
REFERENCE	Wakayama,M.
AUTHORS	Unpublished
JOURNAL	4 (bases 1 to 1497)
REFERENCE	Wakayama,M.
AUTHORS	Direct Submission
TITLE	Submitted (13-FEB-1995) Mamoru Wakayama, Oita University, Applied Chemistry, Dannoharu, Oita, Oita 870-11, Japan
JOURNAL	(Tel:0975-69-3311(ex.746), Fax:0975-69-7957)
FEATURES	Location/Qualifiers
SOURCE	1..1497
	/organism="Achromobacter xylosoxydans"
	/isolate="A-6"
	/sub_species="xylosoxydans"
	/db_xref="taxon:85698"
	1..1497
	/codon_start=1
	/transl_table=11

/product="N-acyl-D-amino acid amidohydrolase"  
 /protein\_id="BA008350.1"  
 /db\_xref="GI:1800111"  
 /translation="MTDSSTLDPAQADFTIAGTITLDGGGPARQDGLAVGRIV  
 ALGDRHAPFVIDARGLALAPFTISHTHDDGLLAHPELIPVSGITVTYVNC  
 GISTAPLSRQIPQDLGLPEPELRFATFDMRLALAPPAVAVIPLVHTTLRA  
 VMDTGRATADERAMRALDLEALQACFVSTGTFPPASAPATDEIDVOPLRG  
 RAGAIYATHREDADHIVPAMEALIGRELDRCRVFHHKLAGERNRGRSETLDM  
 SRAATOVICDCHFPATSTMLRLDRALRSRLITWSKCYPEATGDSFVMAELG  
 LDDEAATARIAPAGAIYFLDMQADVNRIFSHPLTVGSDLPDFHPRMGITNV  
 LRTNVRQRLSLETAHKMTGLAAQUGLEERGLKQGLHADVLEHPIFANTDTATF  
 SAPIQVSGIAHVNWNGROYMDERTGAERPOQVLAIPDAIPWQOSE"

Query Match 39.5%; Score 693.6; DB 1; Length 1497;  
 Best Local Similarity 67.7%; Pred. No. 6,3e-67;  
 Matches 988; Conservative 0; Mismatches 469; Indels 3; Gaps 1;

45 CGATTCCAGCCCTTCTGCTGCTGCGGGGCGACCTTCATGAGCGACAGAAC 104  
 27 CGGTCCCGGCAAGCGCATTCATATCCCGGCCACCTCATGAGCGGCGCGG 86  
 105 CCCGGGGCGCGCGGACCTGGGCGTGGCGGCGACCGCATGCCCATCGCGCAT 164  
 87 CCCCGCCCGCAAGGCGACCTGGGCGTGGCGGCGGCGCATCGTGGCGCATTT 146  
 165 GTCGAGCGCGCGCGCGACACCGGGTGCAGTGTGCGGGCGTGTGTGGCCCGCGCTT 224  
 147 CGCCCATGCGCGCGCGGTGCGGTGATGATGCGCGCGCGTGGCGCGTGGCGCTT 206  
 225 CATGCACTGCACACCGACAGACAACTGCTGTGACGCGTGGAGATGACGCCAA 284  
 207 CATGCACTGCACACCGACAGAGCGTACGCTGTGGGCGATCCGAGATGTGCCAA 266  
 285 GATCTGCGAGGCGTGTACACAGGTGTGTGAGGCAATTCGGGATGACCTGGCGCGCT 344  
 267 GGTTCGCAAGGCGTGTACACAGGTGTGTGAGGCAATTCGGGATGACCTGGCGCGCT 326  
 345 GCGCGACGCGCAACCGCGCGCGCGCGTGTGAGCGTGTGAGCGAAGGCGCTTACCGTTT 404  
 327 GTCGCGCGCGGAGATCCCGACGCGTGTGACCTGTGTGGCGCGCGCGTGTTCGTTT 386  
 405 CGAGCGCTTGGCGGACCTGTGAGCGGCTGTGGCGCGCGCGCGCGTGTACGCGCGC 464  
 387 CGCCACCTTTCGCGGACTGTGGCGCGCGTGTGGCGAAGCGCGCGCGTGTACGCGTGT 446  
 465 CTGTATGCTGGCGCATTCACGCTGTGGCGCGCGCGTGTGTGAGCGTGTACGCGCGCGC 524  
 447 CCGCGTGTGGCGCGACACCGTGTGGCGCGTGTGTGAGCGACACCGCGCGCGCGC 506  
 525 CACCGAGAGAAATGCGCGCATGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 584  
 507 CACCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 566  
 585 CATCGGCAATTTGACGCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 644  
 567 CTTCGCGCGTGTACCG 626  
 645 CATCGGAGTGTGCG 701  
 627 CATCGAGCTGTGCG 686  
 702 CGACGAGGCGGAGCATGTGCGCGCGCGTGGAGGAACTTCGCGCGCGCGCGCGAGCT 761  
 687 CCACGAGGCGGAGCATGTGCGCGCGCGCATGGAAGGCGCGCGCGCGCGCGCGAGCT 746  
 762 GGACGTGCGGT 821  
 747 GGACGTGCGGT 806  
 822 GCGCGAGGCGGT 881

Db 807 CCGCGAAGCGTGTGAGCATGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 866  
 QY 882 GATCCCTTACGTGGCGCGCGTCCACCATGCTCAAGAGACCGCGTGTGGCGGACG 941  
 Db 867 CCACCGCTTACCG 926  
 QY 942 CACCATCATCAGCTGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1001  
 Db 927 CACCGCTTACCG 986  
 QY 1002 CCG 1061  
 Db 987 CATGGCGGAACTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1046  
 QY 1062 CATCTACTTATGATGAGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1121  
 Db 1047 CATCTACTTATGATGAGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1106  
 QY 1122 GATCGGCTCGAGCG 1181  
 Db 1107 GGTCCGCTGCGAGCG 1166  
 QY 1182 CCG 1241  
 Db 1167 CACCAAGCGTGTGCG 1226  
 QY 1242 ATGGAAGATGACCG 1301  
 Db 1227 CCACAAGATGACCG 1286  
 QY 1302 GCGCGGCTCTTCCG 1361  
 Db 1287 GAGGCGGTATGATGCTGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1346  
 QY 1362 CTTCGAACCGCTTACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1421  
 Db 1347 TTTTTCAGCTTCCATTCAGTGTGACGAGGCGCATCCAGCGCGCGTGTGTGTGTGT 1406  
 QY 1422 GGTCTGGCAAGACGAGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1481  
 Db 1407 GGTCTGGGATGGGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1466  
 QY 1482 CCGCTGAGCG 1501  
 Db 1467 TGCATTAACCTTGGAGTCAAC 1486

RESULT 8  
 AL646083  
 LOCUS  
 DEFINITION  
 Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;  
 AL646083 AL646053  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Ralstonia solanacearum.  
 Ralstonia solanacearum  
 Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 Ralstonia.

REFERENCE  
 1 (bases 1 to 208050)  
 Salanoubat,M., Genin,S., Artiguenave,F., Guzy,J., Mangenot,S.,  
 Ariat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,  
 Chandler,M., Choisme,N., Claudel-Renard,C., Cunne,S., Demange,N.,  
 Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,M., Schlex,T.,  
 Sigler,P., Thebaud,P., Whalen,M., Winkler,P., Levy,M.,  
 Weissenbach,J. and Boucher,C.A.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

2 (bases 1 to 208050)  
 Direct Submission  
 Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
 Creteil, Cps706, 91057 Evry Cedex, France, Laboratoire de Biologie

COMMENT  
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS  
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA  
UGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,  
laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,  
F31326 Castanet-Tolosan Cedex, laboratoire de Genetique Cellulaire  
INRA, BP27, F31326 Castanet-Tolosan Cedex  
Christian.Boucher@toulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.

FEATURES  
Location/Qualifiers

1. 208050  
/organism="Rastonia solanacearum"  
/strain="GM11000"  
/db\_xref="taxon:305"  
/plasmid="megaplasmid"  
135..947  
/gene="RSp1115"  
/note="RS05505"  
135..947  
/gene="RSp1115"  
/note="RS05505"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
Predicted by Codon\_usage  
Predicted by Homology  
Predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CAD18266.1"  
/db\_xref="GI:17431589"  
/translation="MADSVYOQVADVPGVLVLSGRFSQSFERRHFLDPHVLAVTAG  
VORRFKGVYLVGPGCISLMPGEIHGVAEAGAYTKFRLSOALLNADISG  
CHREPERAATLEDPALVAGHLRLHDMOMSGASTLOVOSPLALFILKOSALLP  
EHRGALSAPQOMORVRYCFRAGERTITLDSLCLGRFOFKOKOTYGMPPHAM  
LVRLTQACSLSSRSKALAEVQEVFTDOSHFRNFRQAFGVAPSHYS"  
1162..2259  
/gene="RSp1116"  
/note="RS05504"  
1162..2259  
/gene="RSp1116"  
/EC\_number="4.1.2.5"  
/function="small molecule metabolism; degradation; amino  
acids"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
Predicted by Codon\_usage  
Predicted by Homology  
Predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PUTATIVE LOW SPECIFICITY L-THREONINE ALDOOLASE  
PROTEIN"  
/protein\_id="CAD18267.1"  
/db\_xref="GI:17431590"  
/translation="MNRDPMPTTRPALGFTSDNIEGASPEVLDAIILANAGAPYG  
ADCTARVEORTLSEFEREVKVLVPTGTANALCIGAMPMPGNYVCPSHINDE  
CGAPEFTNGAKLVAVDGPRAKIDYERKRAVTRKYGVSHTOPACVSLTATEYSEL  
YTLNEITAIIGQLCKDARKIKLOMDSRFANALVSGCSAANTWAGVALLSFGATKNG  
VVAEAVILFDLSLAEMGYRRKRAGHLSFMRFLSAOIEAYLANDLMLRNARQANDM  
ALRLARLEGHAGSALVGAETANIVYCHLPSTLLENLLQAGFRFYHNMEPVNVEFT  
SEATYADVDVNLRLHMKAAAGPSADTATA"  
2257..5019  
/note="RACR or RS05503  
Predicted by Framed"  
/evidence=not\_experimental  
complement(2462..3139)  
/gene="RSp1117"  
/note="RS05502"  
complement(2462..3139)

/gene="RSp1117"  
/function="miscellaneous; unknown"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
Predicted by Codon\_usage  
Predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PUTATIVE TRANSMEMBRANE PROTEIN"  
/protein\_id="CAD18268.1"  
/db\_xref="GI:17431591"  
/translation="MNDASLHETISCFLEHORPPVRIAEFRNDEAQNVALRAL  
ADNHGVLLPHSDVNIAPHFSAPPTCVKGNRKMWCNACALGLVHLGASATI  
EVRIGAIIDHVIIRIENGRLDLDVYHPIPKHAMDVITYCSVQLFRDAVDYD  
WCSRGIPKGDVPIEQVNNFAVEYARHADADWTKWSVREAIELFARHHLSPVWTI  
AAEARF"  
complement(3188..3595)  
/gene="RSp1118"  
/note="RS05501"  
complement(3188..3595)  
/gene="RSp1118"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
Predicted by Codon\_usage  
Predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAD18269.1"  
/db\_xref="GI:17431592"  
/translation="MFDTKIAYVVDRLATWOKLNTYAFILSGITGASPEINGEAYRD  
AANTYHSLCQGVIVISGDEFTLKAHPRALERKKPHGIXIEDMSSTGHDAANRAVF  
AENGPDSARVYGLAREKTIYDKRGARMP"  
complement(3673..4161)  
/gene="RSp1119"  
/note="RS05500"  
complement(3673..4161)  
/gene="RSp1119"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
Predicted by Codon\_usage  
Predicted by Homology  
Predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAD18270.1"  
/db\_xref="GI:17431593"  
/translation="MGRSEYGETVCPARTLAVYGDWYVLIIRLEFLGYKREERQ  
QTGMSHLISLTRLKLEADGIVVRLHLSDBPPRHERLTSKGDFTPLLSLSWSEK  
WGFKKAKTAPALYITTHROCGHETGLKLVCPACDEPGPKDATVTTLGASFAERQARD  
GA"  
4228..4725  
/gene="RSp1120"  
/note="RS05499"  
4228..4725  
/gene="RSp1120"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
Predicted by Codon\_usage  
Predicted by Homology  
Predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"



gene

CDS

gene

/protein\_id="CAD18271.1"  
/db\_xref="GI:17431594."  
/translation="MPATIRHPRCAPRYLEDYFVGDLTTETSGFAVTRDMILSFAEYOY  
DPQPHMDIARAEIVTGFELVSGWOTLAVTMRLLDRLALGTPIGCAELRDIIFRAA  
PMRPEDVITARAEEVAIAIRPSKSRPDGMADARVTTTNSEGVTLTVQRMSLVPRRPE  
OKVAF"

complement(5017..5610)  
/gene="RSP1121"  
/note="RS05498"  
complement(5017..5610)  
/gene="RSP1121"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : putative  
gene name confidence :  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence-not\_experimental  
/transl\_table=1  
/product="PUTATIVE "TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CAD18272.1"  
/db\_xref="GI:17431595."  
/translation="MAEKITRKPRADAERNRRQLDVAKTAERKVSASLEIDIAE  
AGVGIGTIRHPTFDALIDEIYDESSRLAEAROSEDPLEAVRWMLLFVGL  
ANKAIRADVLCNMGDSERCLTSGEVLIEITLADIIRAROSGAIGLAVPDLISNV  
AGVASFGIELMDCGARLVEVMVAAGLRVCATGGR"  
5822..6910  
/gene="RSP1123"  
/note="RS05497"

Query Match	30.4%;	Score 533.6;	DB 1;	Length 208050;
Best Local Similarity	61.3%;	Pred. NO. 2.9e-50;		
Matches 876;	Conservative 0;	Mismatches 549;	Indels 3;	Gaps 1.

OY	57	CTGCACCGTGTGCTCGGGGGGGGCGACACCTCATTCGACGGCAGCAACACCCGGGGCGGCT	116
Db	95887	CATCCGACAGGGTCCGACATGTGTGAGCGGCACGGCCGACGAGCCGGATGCACACTTTCCGA	95946
OY	117	CGCCGACCTGTGGGCTGTGGCGGCGACCGCATTCGCGCAGTCGGGCAATCGGCGATGTCGGAGGGCGC	176
Db	95947	CGTCCCATTCGGCGGATGGCCCGCATGGCCGCATTCGCCACCTTCGATCCACCGCTGGCT	96006
OY	177	CGCGCACACCCGGGTGACGTGTCTGGGCTGTGGTGTCTGCGCCCGCCGCTTCATCGATTCGA	236
Db	96007	GGGCGACGAGGAGTGTGAAGGGCGGGCGCGGTGTGGGGCCCGGCTTCATGACGTCCA	96066
OY	237	CACCCAGCAGCAACTACTGTCTAGCGCTGCGCATATGACGCGCCAGATCTCGCAGG	296
Db	96067	CACCCAGCAGCAGCAACTGTATCCGACGCGCGGACATGTCTCGGAGAGTGTCCAGG	96126
OY	297	CGTGCACAGGGGGTCAAGGGGCAATTGGGGCATATAGCCTGGGCGCGCTGGCGCAC---GC	353
Db	96127	CATTCACACCGGTATCTGTGTGGCAATGTGGGCAATCGACGATCGCGCGGACGCTCCGGGG	96186
OY	354	CAACCCGCCCCCCCCCTGGACCTCTGAGACAAAGGGGGCTTACCGTTTCGACCGCTT	413
Db	96187	CGAGCGCGCGGAGCCCGATGAACCTCTCTGGCCCGCGCGGGCTTCCGTATCCACCTT	96246
OY	414	CGCGCACTACTGTGACGCGGTTCGGGGCCACGCGCGGCGCGCTGCATACGCGCGCTGTATGGT	473
Db	96247	CGCGTCGTACGTGAGGGGGGTGGAGACAAAGCGCGCGCGCTGCATACGTGGCGGCGCTGT	96306
OY	474	GGGGCATTCAACGCTGTGGCGCGCGCGGTATGCGCGGACTTTCAGCGCGCGCCACCGAGA	533
Db	96307	CGGCGACACGCGCGCTGCGCCACAAACACACTGCAGACGGCTCGACCGCCCGCCACAGGGCGC	96366
OY	534	GGAAATCGCGGCGCATGCGGGGAGCTGTGGCCGAGAGAACCATGGCCAGCGGGCATGTGGCAT	593
Db	96367	CGAAGTCAACACCATGCGCGCGGCACTCGGAGAGCGTCTGCACACCGCGGCGCTGGGCTT	96426
OY	594	TTTCGACCGGGCGCTTCTATCCCGCGCGCGCGCGGCGGCGGCGGACACCGGAAGATCATGTGAGT	653

REFERENCE	ORGANISM	LOCUS	DEFINITION	FEATURES	SEQUENCE	COORDINATES
96427	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
654	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
96487	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
714	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
96547	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
774	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
96607	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
834	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
96667	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
894	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
96727	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
954	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
96787	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
1014	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
96847	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
1074	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
96907	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
1134	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
96967	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
1194	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
97037	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
1254	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
97087	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
1314	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
97147	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
1374	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
97207	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
1434	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486
97267	<i>Staphylococcus aureus</i>	SA000000	Genome		ATGCTGATGACCT	96486

**TITLE** Studies on structure and function of bacterial N-acyl-D-amino acids  
**JOURNAL** amido hydrolyase  
**REFERENCE** Thesis (1995) \_FACULTY\_/Applied Chemistry/\_MAJOR\_, Oita University  
**AUTHORS** 2 (bases 1 to 1467)  
**TITLE** Makayama, M.  
**JOURNAL** Direct Submission  
**REFERENCE** Submitted (06-MAR-1995) Mamoru Makayama, Oita University, Applied  
**AUTHORS** Chemistry; Danno, Oita, Oita 870-11, Japan  
**TITLE** (tel:0975-69-3111(ex.746), Fax:0975-69-7957)  
**FEATURES** Location/Qualifiers  
**source** 1. 1467  
 /organism="Achromobacter xylosoxidans"  
 /strain="A-6"  
 /db\_xref="taxon:85698"  
 1. 1467  
 /codon\_start=1  
 /transl\_table=1  
 /product="N-acyl-D-amino acids amido hydrolyase"  
 /protein\_id="BA08779.1"  
 /db\_xref="GI:1707291"

**CDS**  
 /translation="MOERLDVIEGWNIDLGSPRRRVDYINGEIRIAIGDLSAAP  
 ADRLDAGRIYAFEDIDHMDLMEVEKGLMKTSOGITSVVNGCISGAPAPL  
 PGNRAALALLGDSPLFADMAFYGALEAPRMINVALVGHANRLAAMDPAAPL  
 AKBOBAMERMLADALFAGAVGFSGLAYOPGSAVDOALDGLAVAAARGALHSHIR  
 NEGBAYEPAUDEVILAVGRRTCATVLSHKKMPANMCKSAATLADRAGVDVA  
 LDITPESSTILIPERADQIDITRTITWSTHPECGGSLAIAAMCCDVTARL  
 CPAGAIYFAMDENVRIRFOHECCVSGDGLPNAPHPRLMGSFTRLVIGRYVEAEI  
 LTLBAAYAKTALPARVFLADRLAGLAAGADVAVFADVYCDRATWDAPTLASAGI  
 EHVLVNCAVYPOAPPSHPRILRLRDSINAGAEFFSR"  
 BASE COUNT 213 a 530 c 530 g 194 t  
 ORIGIN

Query Match 25.9%; Score 455.8; DB 1; Length 1467;  
 Best Local Similarity 59.3%; Pred. No. 4.8e-41;  
 Matches 813; Conservative 0; Mismatches 552; Indels 6; Gaps 2;

59 TCAGCTCTGCTCCGCGGCGGACCCCTCATGAGGCGCAACCCGCGCGCGG 118  
 14 TGGATCTGTCATCAAGCGGCTGGTATGACGCGCTGGCGCGCGCGCGG 73  
 119 CCGACCTGGCGCTGGCGGCGGACCCCTCATGAGGCGCAACCCGCGCGG 178  
 74 CCGAGCTGGCGATCGCTGGCGGCGGACCCCTCATGAGGCGCGCGCGG 133  
 179 CCGACACCGCGGCTGGCGGCTGGCGGCGGCGCGCGCGCTTCATGATG 238  
 134 CCGAGCGGCGCTGGCGGCGGCGGCGGCGCGCGCGCTTCATGATG 193  
 239 CCGACGAGCAACATGCTGCTGAGGCGGCGGCGGCGGCGGCGGCGG 298  
 194 GCGACGATGACCTGATGTTCTGGAAGAAACCGGCGCTGGAATGGAAG 253  
 299 TCACACGAGTGTCAAGGCGGCAATGCGGATGAGCTGGCGCGCGGCG 358  
 254 TCAGCTGGTGTGTCGCAACTGCGGATGAGCGGCGCGCGCGCGCGG 313  
 359 CCGCGCGCGCGCTGGAGCTGCTGAGAGGAGCGGCTTCATGAGCGCT 418  
 314 ATACCGCGCGCGCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 370  
 419 ACTACCTGAGCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 478  
 371 TGTACTTGGCGCGCTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 430  
 479 ATTCAACGCTGGCGCGCGGCTGATGCGGATGCGGCGGCGGCGGCGG 538  
 431 ACGCAACTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 490  
 539 TCGCGCGCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 598  
 491 ACGCGCGCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 550  
 599 CCGCGCGCTTCAACCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 658

551 CCGGCTGTGGCTTACCAAGCGCGCGGCGGCTGCGCGGCAAGCGCACTGAGCGCTGGCGG 610  
 659 GCGCCCTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 718  
 611 GCGTGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 670  
 719 TCGTGGCGCGCGCTGGAGGAAACCTTCCGATCGGCGGCGGCGGCGGCGG 778  
 671 TCGAGGCGCGCGCTGACGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGG 730  
 779 TCTCCGACCAAGATGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 838  
 731 TCTCCGACCAAGATGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 790  
 839 TGTATGCA---GGCGCGCATGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 895  
 791 ACATGAGCGG 850  
 896 CCGGCTCAACCATGCTCAAGCAGAGCGGCGGCGGCGGCGGCGGCGGCGG 955  
 851 GCACTTCCACCATCTGATATACCGGAGCGGCGGCGGCGGCGGCGGCGG 910  
 956 GGTGCAAGCGCTTCCCGCACTGAGCGGCGGCGGCGGCGGCGGCGGCGG 1015  
 911 GGTGCAAGCGCGATCCGAAATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 970  
 1016 GCAATTCAGAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1075  
 971 GCTGCGAGCGGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1030  
 1076 TGGAGCAACCGCATGCTGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1135  
 1031 TGGAGCAAGAACAGGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1090  
 1136 GCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1195  
 1091 GCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1150  
 1196 GCGATATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1255  
 1151 GCGGCTACGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1210  
 1236 GCGTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1315  
 1211 GCGTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1270  
 1316 CCGACCTGGTGTGTCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1375  
 1271 CCGAGTGTGTGTCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1330  
 1376 CCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1426  
 1331 CCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1381

RESULT 10  
 LMFCHR34\_10/c  
 Sequence split into 18 fragments Locus LMFCHR34 Accession AL499623  

Fragment Name	Begin	End
LMFCHR34_00	1	110000
LMFCHR34_01	100001	210000
LMFCHR34_02	200001	310000
LMFCHR34_03	300001	410000
LMFCHR34_04	400001	510000
LMFCHR34_05	500001	610000
LMFCHR34_06	600001	710000
LMFCHR34_07	700001	810000
LMFCHR34_08	800001	910000
LMFCHR34_09	900001	1010000
LMFCHR34_10	1000001	1110000
LMFCHR34_11	1100001	1210000

LMFLCHR34.12 1200001 1310000  
 LMFLCHR34.13 1300001 1410000  
 LMFLCHR34.14 1400001 1510000  
 LMFLCHR34.15 1500001 1610000  
 LMFLCHR34.16 1600001 1710000  
 LMFLCHR34.17 1700001 1720777  
 Continuation (11 of 18) of LMFLCHR34 from base 1000001 (AL499623 Leishmania major chromo

Query Match 17.4%; Score 306; DB 2; Length 110000;  
 Best Local Similarity 49.2%; Pred. No. 2.1e-25;  
 Matches 489; Conservative 0; Mismatches 505; Indels 0; Gaps 0;

QY 650 AGGTGTGGCGGGCGGTGAGCGCGCATGGCGCATATCCGCCACACATGCCGCGACAG 709  
 DB 92009 AAGNN  
 QY 710 GCGACACATCGTGGCGCGCGTGGAGAAACCTTCCGCATCGCGCGAGCTGAGCTGC 769  
 DB 91949 NNN  
 QY 770 CGGTGTATCTCGCACACACAGTCAATGGCCACGCCAATTTCCGCGCTCGCGGAGA 829  
 DB 91889 CGACCGTATCGCATACCAAGCTGTGTGGCGAACGCAATCAGCGCGCGACCGGAGA 91830  
 QY 830 CGCTGCGCTGATCGAGCGCGCATGGCGCGCGCGAGAGCTCGCTGGAGCGCGTATCCCT 889  
 DB 91829 CGCTGGCGCGAGTGGGGAGCAATGGCGAGAGAGATGCCCTTGTGCTGAGCTGATCCGT 91770  
 QY 880 ACGTGGCGCGCTCGACCATGCTCAAGAGAGACGCGCTGCTGGCGCGAGCGACCATCA 949  
 DB 91769 ACGCGCGCTCGTGCACATGCTGCTCGCGAAGCGGTGCGACAGTGGCGCGCATCTGA 91710  
 QY 950 TCACCTGGCGCAAGCGCTTCCCGCAACAGCGCGCGCGAGCTGATGAAGTCCGCGCG 1009  
 DB 91709 TCACCTGGCGCGCGCGCATCGGAGAGCGCGCGCGCTGATGAGAGTTCGCGCGAC 91650  
 QY 1010 AGCGCGCGCAATCAAGTACAGAGCTGTGTCGCGAGCTGACGCGCGCGCGCATCTACT 1069  
 DB 91649 AATGGCGCGCGCTCGCGCGCGCGAGTGTGCGCGAGCGCTTGTGCTGCGCGCGCGGTACT 91550  
 QY 1070 TCATGATGAGAGACCGACAGCTGAGCGCGCTGCGCGTTCGCGCGCGACCATGATCGGT 1129  
 DB 91589 TCATGATGAGAGAGCGCGCATGTGCGCGAGCATCTTAAGTACCGGAGACAGATGATCGGT 91530  
 QY 1130 CGAGCGCGCTGCGCGACAGAGCGCGCGCGCGCTGCGCGCGCGCGCATCTTCCGCGCG 1189  
 DB 91529 CGAGCGCGATTCGCTCGCGACAGAGCGCGCGCATCCCGCGCTGCGCGCGCGCGCG 91470  
 QY 1190 TGCCTGGCGACATGCGCGCGAGCTGCGCTGTCGCGTGGAGAGCGCGGTATGAGA 1249  
 DB 91469 TGCCTGGAGTGTACTGCGCGCGAGTCTCGCGCTTTCGCTGCGAGCGCGCGGTTCACAGA 91410  
 QY 1230 TGACGCGCTGACCGCGCGCGCGCTTGGCGCTGCGCGCGCGCGAGCTGACGCGCGGT 1309  
 DB 91409 TGACGCGCTGCGCGCGCGCGCGCTTGGCGCTGAGAGCGCGCGCGAGCTTGGCGGGA 91350  
 QY 1310 ACTTGGCGCGACTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1369  
 DB 91349 AGTGTGGCGCGACTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 91290  
 QY 1370 ACCCTACGAGCGCGCGCGCGCGCGCATTCCTGTAGTCAAGCGCGCGCGCGGTGCGC 1429  
 DB 91289 CGCCAGACGCGCGCGCGCGCGCGCATCCGCGAGCTGTCTTCCGCGCGCGAGCGCGTGC 91230  
 QY 1430 AAGACGAGCGCTTCAACCGCGCGAGATGCGCGCGCGCTGCTGCGACGAGCGCGCGCTGAG 1489  
 DB 91229 GCGATGCGGTGCTACCGCGCGCGCGCGCGCGCTGCTGCGACGCGCGCGCGCGCGCG 91170  
 QY 1490 CCGCGCGCGAGCGCTTCAATCGCGCGGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1549  
 DB 91169 NNN 91110  
 QY 1550 TGAGCGCAACCGCTACATGCGCGCGCTCCGCTCGCAATACGCGCGCGCGCGCGATTCG 1609

DB 91109 NNN  
 QY 1610 TGGCGAAGAGTATGCG 1643  
 DB 91049 TGGCGTGGCG 91016

RESULT 11  
 SC2K36/c  
 LOCUS 22251 bp DNA linear BCT 12-MAY-2002  
 DEFINITION Streptomyces coelicolor cosmid 2K36.  
 ACCESSION AL591857 AL645862  
 VERSION AL591857.2 GI:20520740  
 KEYWORDS aminotransferase; bifunctional protein; carbohydrate kinase; D-amino acid deaminase; D-aminocyclase; iclR-family transcriptional regulatory protein; metal ion transport protein; phosphoenolpyruvate carboxylase; RNA polymerase ECF sigma factor; sugar transport integral membrane protein.  
 SOURCE ORGANISM  
 Streptomyces coelicolor A3(2).  
 Streptomyces coelicolor A3(2).  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomycetales; Streptomycetaceae; Streptomycetes.  
 1 (bases 1 to 22251)  
 Redenbach, M., Krieser, H.M., Denapalle, D., Eichner, A., Cullum, J., Kinash, H., and Hopwood, D.A.  
 A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
 Mol. Microbiol. 21 (1), 77-96. (1996)  
 97000351  
 8643436  
 2 (bases 1 to 22251)  
 O'Neill, S. and Harris, D.  
 Unpublished  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 JOURNAL  
 TITLE  
 JOURNAL

## COMMENT

Notes:  
 Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web  
 (URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))  
 CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (troph et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/>  
 jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (aag, gta, tta or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a

small overlap between neighbouring submissions. Cosmid 2K36 overlaps cosmid SK15 and cosmid SK13.

location/Qualifiers

1. 22251

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db\_xref="taxon:100226"

/clone="cosmid 2K36"

3. 112

/note="nominal overlap with Streptomyces coelicolor cosmid SK13"

complement(160..864)

/gene="2SCK36.01c"

/note="SCO4978"

complement(160..864)

/gene="2SCK36.01c"

/note="2SCK36.01c, conserved hypothetical membrane protein, len: 234 aa; similar to SW:Y485.MYCTU (EMBL:AL021897) Mycobacterium tuberculosis hypothetical 26.0 kDa protein RVI085c or MY017.3bc, 242 aa; fasta scores: opt: 563 Z-score: 653.6 bits: 128.3 E(): 8.6e-29; 39.524% identity in 210 aa overlap. Contains possible hydrophobic membrane spanning regions"

/codon\_start=1

/transl\_table=1

/product="putative integral membrane protein"

/protein\_id="CAC40591.1"

/db\_xref="GI:14285272"

/db\_xref="SPTREMBL:093JL6"

/translation="MTASVPDAPRDLPAGGRGPTLSHPVKKLRLGMLHGFPAV LVAGLVLTALMETSEARIGAGVYVLRCLLGVSAHLHKGMSPRMGCVLRDLHANI FLIAGTYPLTMLLPAGKEMLMGIMAAAGIVRFAVWGPAPLTYPCYLAAG WAAVYLPFEMRAGGAVLAVLYVIGLTVSGAVYGLKRPNSPRMGFEVHSEFT LAAFLAHVVGISLVAY"

1141..1145

1151..2980

/gene="2SCK36.02"

/note="SCO4979"

/gene="2SCK36.02"

/note="2SCK36.02, probable phosphoenolpyruvate carboxylase, len: 609 aa; similar to SW:PECK\_CHLII (EMBL:S56812) Chlorobium limicola phosphoenolpyruvate carboxylase [GTP] (EC 4.1.1.32) PCKA, 646 aa; fasta scores: opt: 2400 Z-score: 2689.0 bits: 507.7 E(): 3.6e-142; 58.361% identity in 598 aa overlap. Contains Pfam match to entry PF00821 PEPCCK, Phosphoenolpyruvate carboxylase and match to Prosite entry PS00505 phosphoenolpyruvate carboxylase (GTP) signature"

/codon\_start=1

/transl\_table=1

/product="putative phosphoenolpyruvate carboxylase"

/protein\_id="CAC40592.1"

/db\_xref="GI:14285273"

/db\_xref="SPTREMBL:093JL5"

/translation="HARDIAPPPVTAAADPDVAVEDRTFCSAKEDAGTNNMKD RLGGELVKGTRKLDPIKRPNSYTAASDPDVAVEDRTFCSAKEDAGTNNMKD PAEMRAITGDEGEGELFRGSMRGRTYVVPCKPLSGPLSALEVTSHSYVASM RMTNRGQVLDDELDEGFEVAVSVAGVAPGQADVPNCSTKYETSHSPEDIEIN SVSGYGNALIGKKCYALRIASVADRGEMLAEMLVKLTPPGARKYVAAAPESA CCKTLAMLEPTISGWYETIGDDIANKRGEGDLYAIINPAGFPVAPGGEPTNA NAKTTLNGSVFTNVALTDGQVWMEGTETEPALHTDKGNDTPEGTPAHPNAR FTTPAACPIIAPEMEDPGVPISAILFGGRATVPLVTESFDNHHGFLGANVASE KTAALAGKVGELRDPFAMLPFGYNGMGYGMHNDVAKDKQSKLPKTYVYNNFRKD DAGRYVMPGFGENGKLVIVERLEGRAADGETPIGVLPTEKSLDTGDLADADLEF LLSVQKVVREERAAALVPEHLNFDGHTPAELMDQYRALVRRUG"

1205..2974

/gene="2SCK36.02"

/note="Pfam match to entry PF00821 PEPCCK, phosphoenolpyruvate carboxylase, score 1344.70, E-value 0"

/gene="2SCK36.02"

misc\_feature

1967..1993

/gene="2SCK36.02"

/note="PS00505 Phosphoenolpyruvate carboxylase (GTP) signature"

complement(3263..4213)

/gene="2SCK36.03c"

/note="SCO4980"

complement(3263..4213)

/gene="2SCK36.03c"

/note="2SCK36.03c, unknown, len: 316 aa"

/codon\_start=1

/transl\_table=1

/product="hypothetical protein 2SCK36.03c"

/protein\_id="CAC40593.1"

/db\_xref="GI:14285274"

/db\_xref="SPTREMBL:093JL4"

/translation="MSGDGLEGLLAAVHRLTAQRMGPVCGTAGHGTGAVCLVPAGA GLPCLIEAIGSRFEGPLRLVGTTEAAGSAGIPLTPGGERIVEMRAVADRMIG CGYVRADAGAVPVLVAVRADAPAGAGATPVVAGIAYVGMETARAVADMPAV EARLGTPLDGRVRLVIFPGAGAFDGLQVDPADARSTQDVRHREMLGEMARNG SRLMEPRYPVAPBGLLQMASTROADPFYTLTGDPDRPVLAKKEVPSMNERFQPTGFEVHMLTEPDHPFSTAKMFDVHMVDYPPAADGRV"

complement(4210..5610)

/gene="2SCK36.04c"

/note="SCO4981"

complement(4210..5610)

/gene="2SCK36.04c, possible bifunctional protein, len: 466 aa; N-terminal region presents low similarity to N-terminal region of TR:O14362 (EMBL:Z97992) Schizosaccharomyces pombe putative glucan synthase-1 (EC 2.4.1.34) SPBC30D10.17c, 504 aa; fasta scores: opt: 176 Z-score: 188.3 bits: 44.3 E(): 0.0071; 24.931% identity in 361 aa overlap and C-terminal region to C-terminal region of TR:AAK22951 (EMBL:AE005774) Caulobacter crescentus oxidoreductase, GFO/IDH/MOCA family CC0967, 307 aa; fasta scores: opt: 146 Z-score: 160.2 bits: 38.4 E(): 0.26; 26.590% identity in 173 aa overlap"

/codon\_start=1

/transl\_table=1

/product="putative bifunctional protein"

/protein\_id="CAC40594.1"

/db\_xref="GI:14285275"

/db\_xref="SPTREMBL:093JL3"

/translation="MSTHDPATWEPRLRLVRAHPERLAGRGHVMGQIGLGMSP VRAHPAGRSALVEDMODERTAVEAVALRAQKRSVSWEATRVGRDALVAVDS GRVPEHGVSPVGTLLVVEGAVEPERRLEAVPAGLPAESAPALLERTLRRLPD AVGTAEKIAEAQTRILGVTLPDEKALRYRVRAQMDQMRGYAAQERVVAAGVELLP LDOXYVADARSPCLMOPATDAVTAVDAAVQGLVSGPGLIAGDNGGDRLLDLTP PGSGHNGOVVMIGHEDSIGGLAESLTVMVNGRAEMHGRMDRPPVAVARLVNIG DDVAVARPELEVYVSGRRECEPRLAPLAPRLTLNHCRLADPLIAGTRRE ITRGPEMERVLVDAGAVPNSLAAHLEVRKEHHPXILANELLSLMDRPLISRTV LEGDLTDTRAGGAR"

complement(5618..5621)

5786..5792

5801..5983

/gene="2SCK36.05"

/note="SCO4982"

/gene="2SCK36.05"

/note="2SCK36.05, putative membrane protein, len: 60 aa, contains possible hydrophobic membrane spanning region"

/codon\_start=1

/transl\_table=1

/product="putative membrane protein"

/protein\_id="CAC40595.1"

Query Match 9.8% Score 171.8; DB 1; Length 22251; Best Local Similarity 49.5%; Pred No. 1.5e-10; Matches 813; Conservative 0; Mismatches 647; Indels 183; Gaps 7;

QY 32 CCATGTCCCAATCCGATTCACAGCCCTTGACCTGCTGCGGCGGACACCTCATCG 91

DB 10533 CCGGACCCAGAGAGAGTCAAGCATGAGAGAGCTGATCATCGAGGCCGCGCTGCG 10474

[illegible]

```
gene
CDS
    <1..1678
    /gene="asm37"
    <1..1678
    /gene="asm37"
    /codon_start=2
    /transl_table=11
    /product="unknown"
    /protein_id="AA054115.1"
    /db_xref="GI:21449383"
    /translation="IVRGVYVLAATGCVLAAPASAAITSPVYSTATVTEVGRGT
    PARSRYVVDTRSNATGYAOGGIGADHLNGLSIRHSGGMATQVRYVPFYSASGY
    ARMTVERADLDGEQASPNPDVYVQVEAPVETSAKVHTYTAGAGSVAQPSANAPN
    PERGLYHTGDCDCADLSADLARVREOGITLAMCVFLAEHRSSPIAQTALDROR
    OFDARVAGVKLVLRFAVYTAASEGDEPAKERVLAIDQLAPLRAANDYAVVQAGFV
    GAGEWYTKRNGAGVANSANDANRKAAYDKLITLAPADRAVOLATPKFKRALYGGG
    STRIGHNDCEFLASDTPGTYEDPAATRYPLEATRTLPWGGECATNPPRTCCPAT
    AELSRFWYTLNTDYQDDVLRAASQCLFVYAKLGYRTLRDSVPTTTTGAALP
    VRNLTNDGYAAAANQRYNLVLRNTTGTALTRLPDADPRTMLPGARALADQIIP
    SLATGOYALLLELDPALPGRPSEYSIQTANTGLMDARTGNDLKOTITVS"
    1825..2142
    /gene="asm36"
    1825..2142
    /gene="asm36"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AA054114.1"
    /db_xref="GI:21449382"
    /translation="WYTSNVLCAKACGAPRACATARRPFCSTACKOAYRLRKA
    GSASHEDEDLRELALVARKARTLEATPPMSAQATLVGLARVVELAAADAGY
    SGG"
    complement(2129..3307)
    /gene="asm35"
    complement(2129..3307)
    /gene="asm35"
    /codon_start=1
    /transl_table=11
    /product="bicyclomycin resistant protein"
    /protein_id="AA054113.1"
    /db_xref="GI:21449381"
    /translation="MAVGKADSAVGMPLPALLIYALTGTTIEMDSVSPFDIARELG
    VGPRAVQITVTLNPLIGYCLGALCSGRPLSDRGRPRVLLGSSAWMLVGLALPTIG
    LLTAREVQIGASAAAVLVFTIGDRIKGRPLAMGVLVNAASALMTCAKPLIGAL
    NRVAGMRGNVLAVCCTATLSLLVFLFLPETRTARATITARSVLADIRLLGSAVL
    AASLTPSLFAVLVEFIAASPELVTAAGLGVTGALMLLVVASFAPASFGAIIIP
    LLGPGAGVCMGLGAVVGVGLGDLGCPAAYTSGVALFCVFAVCAVPEVDEMSY
    PELAGASASATMGRLALVTLTLTALGALSTGPFAPAAVAAVAVVLLSALSPR
    "
    complement(3450..4064)
    /gene="asm34"
    complement(3450..4064)
    /gene="asm34"
    /codon_start=1
    /transl_table=11
    /product="transcriptional regulator"
    /protein_id="AA054112.1"
    /db_xref="GI:21449380"
    /translation="MARRNPERRALLLDAAEVLAEGEGRGLTFRAVDORAGVPAGTA
    SHYFAGRAITLTCGERVYORLLDOAPLGEETPDGPRDSRVTELMHALVDYSAFPTG
    FLALIEIRLEARNPELREVLTAARADIRFNAEHNAASGLPGDGTVALMLALNML
    IIERLTLPDLTDHQRDLVTALVLRLLLAGORPEEAGSPFDESG"
    4151..4723
    /gene="asm33"
    4151..4723
    /gene="asm33"
    /codon_start=1
    /transl_table=11
    /product="dihydrofolate reductase"
    /protein_id="AA054111.1"
    /db_xref="GI:21449379"
    /translation="MRKLVYFIASIDGVIAEGEGRHDFPTAEPVWEHLRAEYDET
    PTHVRPLVGMPLDTPNRRYDVTLMGRSGYOPALDAGVGSYALHROYVVSRLPADP
    AVALVADPVGIVRELKREDGLDVMVLCGGADLAGOLLPEIDELVYKTYVAVAGGPTL
```

---

```
gene
CDS
    FRAGAPPAFRPVYRAFDHGGVTTYRPA"
    complement(4854..5057)
    /gene="asm32"
    complement(4854..5057)
    /gene="asm32"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AA054110.1"
    /db_xref="GI:21449378"
    /translation="WITQONARIPARRSSGSRNARCVAVMPRQPSATRAQPRS
    SNAQDSRAISRGRDRTSLNSA"
    5415..6461
    /gene="asm31"
    5415..6461
    /gene="asm31"
    /codon_start=1
    /transl_table=11
    /product="RNA polymerase sigma factor"
    /protein_id="AA054109.1"
    /db_xref="GI:21449377"
    /translation="WIGYSADGRRRRPVMGVTAEFTWETDEDIKLAATLRDAEA
    ANADSTRAYIKAGIRVLLTADQVELARVEAGLYAEHLIDLPARGSRPVELR
    RYVRDQRAKDHLENLVLSIAKRYTGKMAILDYIOGNTGLVAVAEFDYAKG
    FKSTYATWIKQALITRALDOSRTIRLPVHAEVYTKLARIERDLRNREYATRE
    VAEVLDTPEKVELRQAHARTPVSLDQITGADADTSGDLIDENAVALEVSSALL
    SADLRVLELTLPREAGVYRMRYGPDVGLTIDEVGRVYGTREIRIQIESRTAKL
    OHPARSGLEGYLD"
    complement(6501..9518)
    /gene="asm30"
    complement(6501..9518)
    /gene="asm30"
    /codon_start=1
    /transl_table=11
    /product="cytochrome P450"
    /protein_id="AA054108.1"
    /db_xref="GI:21449376"
    /translation="WVATGTRIPGKPLPLVGNLDLVLTSDLDTPVDFLDRCHREHG
    IVALTFAGQOVFASSSHELVAKMSDPSGKGVHAPLEVOYKOPFAGGLETAGDEWN
    GKARLRLPAFGPTAMRDHFPAMDIAEOMLRMRFGPDHRIYDAVADTLRLPTIA
    LCAFGANFNFYRDRAHFEVDARVSLVAGEARELPVQFFLVRGNQRYDDIATM
    NRIADGIVARALPAGERPDLLERMLCADPVTGERLSARVRYOLATFLIAGET
    TSGLSFVARRILAHPEVLAKADAVDYLGRVPAFEDLARLDYLGVALYRERLHP
    TAPAFALAPPEPELISGHAGCEPVLMPTLHRDPARWRDPDVPDPERPEDE
    ITPCAMPPEHGKACIGREFALQEAIVLALVLFDFDLADPDRRLTKOTLTLP
    DSVLRARPADRPAGATATVETVPPQVATRHGTPPLHVFSGNSGSEGLARTYG
    DGAARGWATVAPLDLDARVALPASGVPVIVSSYNGCAPDMAHEFRMTLOQPDLSG
    VDYVLGCGNLDMSATYQRYVPTLIDEMAAGARLREAGATDARADPFGDMERYEP
    LMPILSAECGVGEITGPRRVEVSDAAGIGDLASAVYLENLEIRGPDGSKRILE
    LRDPDGTSTRTGYLSTLPQNHDPYRRARVRLGTAEVYVYESSAPGLVPGVAL
    RVDELRLRCVLSLPAGAGVARIAEKPCPEERABLAATGTATLLELRPSCAVD
    LALALELPAARTLXSISSAEOREAEVALTVSVAGVSGYLSRVPRDRAVAGIAS
    PPESEFRPADNTVPVVLIAAGTGIAPFRGLRLARALGEPGALLFCGRPELDDL
    YAEFFALGDMLEVDRAYSRHPDGEVRAVHOHRLAMORDRVRELVDAGAVYLCGDATR
    VCPAVEVIGRIGPCAGMDALRAGGYATDVE"
    9664..10245
    /gene="asm29"
    9664..10245
    /gene="asm29"
    /codon_start=1
    /transl_table=11
    /product="transcriptional regulator"
    /protein_id="AA054107.1"
    /db_xref="GI:21449375"
    /translation="MSAAVRREQLVAAALRYMRKDSIAATYTAICAEDMPHGAHY
    CFRSKQELVTLTGTNTSLDAMPISDPLAVVNVRTMLAAFTATCAQDPEALY
    LGELVTLAOREPELRELPMOHHREYVAKVAHLDRALAEAGFDYSVDRALLAEVVS
    LDGVASMSFSGRDELAKASLDQFAVVFALVTR"
    10393..10857
    /gene="asm28"
    10393..10857
    /gene="asm28"
    /codon_start=1
```





```

VERSION      AL18396.2.1.G1:20520797
KEYWORDS     alpha-amylase; glpBI, 1.4-alpha-glucan branching enzyme; glpP,
              glycogen phosphorylase; integral membrane transport protein; pf2,
              6-phosphofructokinase; pta, phosphate acetyltransferase; secreted
              nucleosidase; sodium:dicarboxylate symporter; transicional
              regulator; trxA2, thioiodoxin; two-component sensor kinase/sensor
              kinase.
SOURCE       Streptomyces coelicolor A3(2).
ORGANISM     Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
              Streptomyces coelicolor A3(2)
REFERENCE    1 (bases 1 to 28751)
AUTHORS      Redenbach,M., Kleser,H.M., Denapate,D., Eichner,A., Cullum,J.,
              Knasahl,H. and Hopwood,D.A.
              A set of ordered cosmids and a detailed genetic and physical map
              for the 8 Mb Streptomyces coelicolor A3(2) chromosome
              Mol. Microbiol. 21 (1), 77-96 (1996)
TITLE        8843436
JOURNAL      2 (bases 1 to 28751)
MEDLINE      97000351
PUBMED       8843436
REFERENCE    2 Oliver,K. and Harris,D.
AUTHORS      Unpublished
JOURNAL      3 (bases 1 to 28751)
AUTHORS      Bentley,S.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
REFERENCE    Direct Submission
AUTHORS      Submitted (03-FEB-2000) Streptomyces coelicolor sequencing project,
              Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Camridge
              CB10 1SA E-mail: barrel@sanger.ac.uk Cosmids supplied by Prof.
              David A. Hopwood, (3) John Innes Centre, Norwich Research Park,
              Colney, Norwich, Norfolk NR4 7UH, UK
              On May 9, 2002 this sequence version replaced gi:6941966.
COMMENT      Notes:
              Streptomyces coelicolor sequencing at The Sanger Centre is funded
              by the BBSRC and Beowulf Genomics
              Details of S. coelicolor sequencing at the Sanger Centre are
              available on the World Wide Web.
              (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
              numbered using the following system eg SC7B7.01c.5C (5.
              coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
              strand).
              The more significant matches with motifs in the PROSITE database
              are also included but some of these may be fortuitous. The length
              in codons is given for each CDS.
              Usually the highest scoring match found by fasta -o is given for
              CDS which show significant similarity to other CDS in the database.
              The position of possible ribosome binding site sequences are given
              where these have been used to deduce the initiation codon. Gene
              prediction is based on positional base preference in codons using a
              specially developed Hidden Markov Model (Krogh et al., Nucleic
              Acids Research, 22(42):4768-4778(1994)) and the Frameplot program
              of Bibb et al., Gene 30:157-66(1984) as implemented at
              http://www.nh.h90.jp/
              jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
              correct initiation codon. Where possible we choose an initiation
              codon (atg, gtg, ttg or (at)) which is preceded by an upstream
              ribosome binding site sequence (optimally 5-13bp before the
              initiation codon). If this cannot be identified we choose the most
              upstream initiation codon.
              IMPORTANT: This sequence MAY NOT be the entire insert of the
              overlapping sections once, or longer, because we arrange for a
              small overlap between neighbouring submissions. Cosmid 6A11 lies
              between and overlaps with cosmids 8F4 and 3D11 on the AseI-B
              genomic restriction fragment.
FEATURES     Location/Qualifiers
              1..28751
               /organism="Streptomyces coelicolor A3(2)"
               /strain="A3(2)"
               /db_xref="taxon:100226"
               /clone="cosmid 6A11"
               2..84
               /note="Nominal overlap with Streptomyces coelicolor cosmid
               ST6A11

```

misc\_feature  
Nominal overlap with Streptomyces coelicolor cosmid 8F4.  
complement(6..872)  
/note="Pfam match to entry PF01515 PFA\_PTB, phosphate  
acetyl/putaryl transferase, score 460.50, E-value  
1.4e-134"

RBS  
complement(1975..1978)  
2103..2107

RBS  
2118..2143

gene  
/gene="SC6A11.02"  
/note="SC05426; pfkA2"  
2118..2143

CDS  
/gene="SC6A11.02"  
/note="SC6A11.02, pfkA2, 6-phosphofructokinase, len: 341  
aa; similar to many e.g. SW:Q59126 (PPF\_AMYE)  
6-phosphofructokinase from Amycolatopsis methanolica (352  
aa) fasta scores: opt: 1274, z-score: 1463.6, E(): 0,  
59.1% identity in 342 aa overlap. Also similar to  
TRNEN:SCB51967 (EMBL:AL109661) Pfk, 6-phosphofructo-  
kinase from Streptomyces coelicolor (342 aa) fasta scores: opt:  
1251, z-score: 1437.4, E(): 0, 56.7% identity in 342 aa  
overlap. Contains two Pfam matches to entry PF00365 PFK,  
phosphofructokinase and Prosite match to PS00433  
phosphofructokinase signature"  
/codon\_start=1  
/transl\_table=11  
/product="6-phosphofructokinase"  
/protein\_id="CAB72402.1"  
/db\_xref="GI:6941968"  
/db\_xref="SWISS-PROT:O9L1L8"  
/translation="MRICVITRAGDCPCPLNIVTSVVRADVNGDVGIFEDGYAGL  
LDGRIRALDNLVSGILRAGGTLISSSLERDLREACENAGDH1QNFGLDILPIRG  
EETLVNARPLSDAGLIVGQVPTIDNDISTDRFGEDTAVGATEAMDKTKTAESH  
QGVNVAEVRGAWIALESGMAAGAHGICLPERFPADLVKVAEERFSGNKFAVH  
CAVEGHAPEGSSMDYGKALDKFCHERFGIGITLAELERLRGEAKPVLIGHVGORG  
GVPTADRVLRTRFGMHAVEAHRGDFGRMTALNGTDVVVPLAEAVTELKTVPKDRM  
DEAESF"

misc\_feature  
2121..2804  
/gene="SC6A11.02"  
/note="Pfam match to entry PF00365 PFK,  
phosphofructokinase, score 278.20, E-value 1.6e-101"  
2883..3017

misc\_feature  
/gene="SC6A11.02"  
/note="Pfam match to entry PF00365 PFK,  
phosphofructokinase, score 52.20, E-value 9.9e-19"  
2913..2959

misc\_feature  
/gene="SC6A11.02"  
/note="PS00433-phosphofructokinase signature"  
complement(3153..3866)  
/gene="SC6A11.03c"  
/note="SC05427"  
complement(3153..3866)  
/gene="SC6A11.03c"  
complement(3153..3866)  
/gene="SC6A11.03c"  
/note="SC6A11.03c, unknown, len: 237 aa; no significant  
database similarities"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SC6A11.03c"  
/protein\_id="CAB72403.1"  
/db\_xref="GI:6941969"

CDS  
/db\_xref="SPTREMBL:Q9L1L7"  
/translation="MMRVRRPSPGSGQSPPEPNALARRLRALNMGEVHAHGIRVS  
YGLPVGSLPIVIMERGTALESGMAGELTAGVMSPELIGRPSLSEHVAAGLADL  
EDVARVGLPIVYLMRESDDMGRTDROSALAVYLCTLPDFTVYGRKKLADLRL  
TSVATTRMGCGVPRILAKPLDRRLVLENTGVYLRHDYGRMAATLSWGCGAPDSAGG  
RDLPLDRIVDHFRVAFVRHP"

gene  
/gene="SC6A11.04c"  
/note="SC05428"  
complement(3978..4868)  
/gene="SC6A11.04c"  
/note="SC6A11.04c, possible integral membrane transport  
protein, len: 296aa; similar to many proposed transport  
proteins e.g. TR:Q9X9Y6 (EMBL:AL096743) putative integral

CDS  
complement(3978..4868)  
/gene="SC6A11.04c"  
/note="SC05428"  
complement(3978..4868)  
/gene="SC6A11.04c"  
/note="SC6A11.04c, possible integral membrane transport  
protein, len: 296aa; similar to many proposed transport  
proteins e.g. TR:Q9X9Y6 (EMBL:AL096743) putative integral



REFERENCE	ORGANISM	KEYWORDS	VERSION	ACCESSION	DEFINITION	LOCUS	RESULT 14
Db 17936	AGGACGAGGAGGACCGGAGGCTCTGTCTCCGGCACAGCATATATGAGACCGGAGGCTTGC			AE006004/c			
Qy 1433	AGCAGGCGTTTCACCGGCCAGCATGCGCGGCCGCTGCTTCGACGACGAGCGCGCTGAGCCC						
Db 17996	CCGAGGCGCCAGGAGGAGGACGACACCGCAGCGCGCGCGGAGGCGCGCGCGACGCCA						
Qy 1493	GGGCGCCAGCCCTTTCATATCCGGCGTGAACGGGGGGGCGCTGCCCGCCCTCCCAACCTGG						
Db 18056	CTTGGGCGTTTCGGGCGACATCATCTGCTGGACGAGGGCGCAGAGACTGTCCC-----CGATGG						
Qy 1553	ACGGAACCGGTACATAGGCGCCCTCCCTCCGCTGCCAATAGCGGCCCGACCGGATATCGTGG						
Db 18110	CTTGGCGGCTCTCATATGCGCGCTGCTGCCCGCCACCGCTCGATGACCTGTGTGGCGACCCGG						
Qy 1613	GCAAGGAAGTATGAGGCGCGCGCTGCGCGCGCGGACGCGAAGGCGCGGAAATGACCTTGC						
Db 18170	CGCAGACCGCGGAGGCGCGCGGTGTCTGGCTCTCTGTTCGAATCTCTCCCGCCCTACGTGC						
Qy 1673	AAGACTGTCTCCAGGCGCAGCGGATCGCGCTTCGACCTGTTCGAAGCGCGAGCTGGGCC						
Db 18230	AGGACCGCTGGAGACACACCGCGCTCGGCGTCACTACCGCACCGCGCGCGAGCTCATG						
Qy 1733	AGATCGC 1739						
Db 18290	ACCTCGC 18296						
REFERENCE	ORGANISM	KEYWORDS	VERSION	ACCESSION	DEFINITION	LOCUS	RESULT 14
Db 17936	AGGACGAGGAGGACCGGAGGCTCTGTCTCCGGCACAGCATATATGAGACCGGAGGCTTGC			AE006004	14710 bp DNA linear BCT 12-JUN-2002		
Qy 1433	AGCAGGCGTTTCACCGGCCAGCATGCGCGGCCGCTGCTTCGACGACGAGCGCGCTGAGCCC			AE006004	Caulobacter crescentus CB15 section 330 of 359 of the complete genome.		
Db 17996	CCGAGGCGCCAGGAGGAGGACGACACCGCAGCGCGCGCGGAGGCGCGCGCGACGCCA			AE006004	AE006004.1 GI:13425161		
Qy 1493	GGGCGCCAGCCCTTTCATATCCGGCGTGAACGGGGGGGCGCTGCCCGCCCTCCCAACCTGG						
Db 18056	CTTGGGCGTTTCGGGCGACATCATCTGCTGGACGAGGGCGCAGAGACTGTCCC-----CGATGG						
Qy 1553	ACGGAACCGGTACATAGGCGCCCTCCCTCCGCTGCCAATAGCGGCCCGACCGGATATCGTGG						
Db 18110	CTTGGCGGCTCTCATATGCGCGCTGCTGCCCGCCACCGCTCGATGACCTGTGTGGCGACCCGG						
Qy 1613	GCAAGGAAGTATGAGGCGCGCGCTGCGCGCGCGGACGCGAAGGCGCGGAAATGACCTTGC						
Db 18170	CGCAGACCGCGGAGGCGCGCGGTGTCTGGCTCTCTGTTCGAATCTCTCCCGCCCTACGTGC						
Qy 1673	AAGACTGTCTCCAGGCGCAGCGGATCGCGCTTCGACCTGTTCGAAGCGCGAGCTGGGCC						
Db 18230	AGGACCGCTGGAGACACACCGCGCTCGGCGTCACTACCGCACCGCGCGCGAGCTCATG						
Qy 1733	AGATCGC 1739						
Db 18290	ACCTCGC 18296						
REFERENCE	ORGANISM	KEYWORDS	VERSION	ACCESSION	DEFINITION	LOCUS	RESULT 14
Db 17936	AGGACGAGGAGGACCGGAGGCTCTGTCTCCGGCACAGCATATATGAGACCGGAGGCTTGC			AE006004	14710 bp DNA linear BCT 12-JUN-2002		
Qy 1433	AGCAGGCGTTTCACCGGCCAGCATGCGCGGCCGCTGCTTCGACGACGAGCGCGCTGAGCCC			AE006004	AE006004.1 GI:13425161		
Db 17996	CCGAGGCGCCAGGAGGAGGACGACACCGCAGCGCGCGCGGAGGCGCGCGCGACGCCA						
Qy 1493	GGGCGCCAGCCCTTTCATATCCGGCGTGAACGGGGGGGCGCTGCCCGCCCTCCCAACCTGG						
Db 18056	CTTGGGCGTTTCGGGCGACATCATCTGCTGGACGAGGGCGCAGAGACTGTCCC-----CGATGG						
Qy 1553	ACGGAACCGGTACATAGGCGCCCTCCCTCCGCTGCCAATAGCGGCCCGACCGGATATCGTGG						
Db 18110	CTTGGCGGCTCTCATATGCGCGCTGCTGCCCGCCACCGCTCGATGACCTGTGTGGCGACCCGG						
Qy 1613	GCAAGGAAGTATGAGGCGCGCGCTGCGCGCGCGGACGCGAAGGCGCGGAAATGACCTTGC						
Db 18170	CGCAGACCGCGGAGGCGCGCGGTGTCTGGCTCTCTGTTCGAATCTCTCCCGCCCTACGTGC						
Qy 1673	AAGACTGTCTCCAGGCGCAGCGGATCGCGCTTCGACCTGTTCGAAGCGCGAGCTGGGCC						
Db 18230	AGGACCGCTGGAGACACACCGCGCTCGGCGTCACTACCGCACCGCGCGCGAGCTCATG						
Qy 1733	AGATCGC 1739						
Db 18290	ACCTCGC 18296						
REFERENCE	ORGANISM	KEYWORDS	VERSION	ACCESSION	DEFINITION	LOCUS	RESULT 14
Db 17936	AGGACGAGGAGGACCGGAGGCTCTGTCTCCGGCACAGCATATATGAGACCGGAGGCTTGC			AE006004	14710 bp DNA linear BCT 12-JUN-2002		
Qy 1433	AGCAGGCGTTTCACCGGCCAGCATGCGCGGCCGCTGCTTCGACGACGAGCGCGCTGAGCCC			AE006004	AE006004.1 GI:13425161		
Db 17996	CCGAGGCGCCAGGAGGAGGACGACACCGCAGCGCGCGCGGAGGCGCGCGCGACGCCA						

gene  
CDS  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK25406.1"  
/db\_xref="GI:13425162"  
/translation="MWLISAHKGCEPSVGFENAVLSSCMWGLYLAAPCAROPDPTRPD  
WGRERFRRRSRSGTRRSRGPAPPAQAAPAAAPAAQAAPDAQAAPPAAPAAEPAPAD  
TYVALVGPPTDPEFTVRLDLVDLKVCKPQVAGDEPAQLVSKSGFKKREOWFALEK  
PYPANLPGSNKNKNCVTVTIDYQAPQAQELANGLHDMAWTWNSPOLRLIRNDQTVG  
SDPFRFVSMDDAAGACRAGLVYRLKKLDETSGVKNFERAQLLYSTSR"  
/complement(1018..1278)  
/gene="CC3445"  
complement(1018..1278)  
/gene="CC3445"  
/note="identified by match to protein family HMM"  
/codon\_start=1  
/transl\_table=11  
/product="ATP synthase F1, epsilon subunit"  
/protein\_id="AAK25407.1"  
/db\_xref="GI:13425163"  
/translation="MAKHESLVAPEBELFSGEVDYQAQAGAEQFGLVLAHNAHPMTT  
LREGVATYKDCATKTVFDIOGFAVDVPEGLITLAHAVEAA"  
complement(1278..1529)  
/gene="CC3446"  
complement(1278..1529)  
/gene="CC3446"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK25408.1"  
/db\_xref="GI:13425164"  
/translation="MTEDADLANFCDCDELAEVSPDANAGLIEHDSGLILNLGEERE  
EGEVLTATYRFCEFCGETEIODEDIEALGVEETFN"  
complement(1529..3148)  
/gene="CC3447"  
complement(1529..3148)  
/gene="CC3447"  
/note="identified by match to protein family HMM"  
/codon\_start=1  
/transl\_table=11  
/product="ATP synthase F1, beta subunit"  
/protein\_id="AAK25409.1"  
/db\_xref="GI:13425165"  
/translation="MAKTPAEKPATAPAKKPPAPAKPAAAPKAAAAKAPAAKAPAKRR  
APKAAAPKPAAGLEGATGRLVQIGAVVDIEFGALPAILNAVEVNTATGCRVLA  
PFAHGHGQSYRAIAMDATGSLVGGVQVRTPGPIRVPGPGLTGRIMVIGPIED  
QGPISDSRTHRDAPSFAEQNTAELVYTGIVYIDMCYTTGGKGLGLGAGVKKR  
LYTMDELINNTAKAIGYTVLGAVERREBSGDLHEHTESNAVVDPRANNSGTEGS  
CALVYGGNMEPPGARARVALTGLSLAEFRDEBGDVLVFDNIERFQAQAEVSALL  
GIPSAVQGPITALEMGNLQERISTKSGTSVQAIYVPADLTDPAPTSFAHLD  
ATTVLSBDIAQAIAPAVDPLDSTSRIMDPITIGEHYTVARVQEVLOQYKAKDLII  
ALICDELSEEDKIVVAAARKIORTLSQPFYABEQFTPCAPFVQLKDTINSFQYVD  
GETDHPGEGAFYWGCPRIEAVAKAKMAAA"  
complement(3196..4071)  
/gene="CC3448"  
complement(3196..4071)  
/gene="CC3448"  
/note="identified by match to protein family HMM"  
/codon\_start=1  
/transl\_table=11  
/product="ATP synthase F1, gamma subunit"  
/protein\_id="AAK25410.1"  
/db\_xref="GI:13425166"  
/translation="NASLKEMRNRISVYATOKITKAMQVYAAAKLRSDQAASAR  
YARLASVIANLAAVSGDGAEPKLLAGCGRDRHLYVVAADRLAGGFTSYIARAR  
AHIDGLIAQKQVRYVYCGKKTATOLARPYACRIVETDLSYRDLTISVAPPLADYV  
TREVYEGEDVTVLFYSRFKSVVQOITPGLOLIPAVVEGEGASPTAVTEPESESE  
IETLEPRMLTVOILSALDDNNAAGFYASQMTAMDNATRNAGDMIKRYTLEYNRSQAO  
IKRELLEITSGAEAV"  
complement(4106..5638)  
/gene="CC3449"

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2003, 07:44:25 ; Search time 87 Seconds  
(without alignments)

8893.123 Million cell updates/sec

Title: US-10-009-782-1

Perfect score: 1758

Sequence: 1 gaattccactgacgcgcga.....ccctgagctacgagaagctt 1758

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 389086 segs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications, NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PC07\_NEM\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEM\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEM\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEM\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEM\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEM\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEM\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86	4.9	88421	9	US-09-976-059-1
2	81.8	4.7	1248	9	US-09-860-846-7
3	81.8	4.7	1248	10	US-09-861-289-7
4	81.8	4.7	13613	9	US-09-860-846-3
5	81.8	4.7	13613	10	US-09-861-289-3
6	73.8	4.2	15872	9	US-09-860-846-1
7	73.8	4.2	15872	10	US-09-861-289-1
8	73.4	4.2	4257	9	US-09-825-288A-1
9	72.4	4.1	3189	10	US-09-815-242-4056
10	72.2	4.1	13842	9	US-09-860-846-30
11	72.2	4.1	13842	10	US-09-861-289-30
12	72.2	4.1	36778	9	US-09-860-846-5
13	72.2	4.1	36778	10	US-09-861-289-5
14	69.6	4.0	11220	9	US-09-860-846-32
15	69.6	4.0	11220	10	US-09-861-289-32
16	67.8	3.9	1014	9	US-09-860-846-13
17	67.8	3.9	1014	10	US-09-861-289-13
18	67.8	3.9	13029	10	US-09-815-242-4052
19	64.2	3.7	2712	10	US-09-748-033-4

20	63.2	3.6	1764	10	US-09-815-242-7984	Sequence 7984, Ap
21	62.2	3.5	804	9	US-09-773-748-2	Sequence 2, Appli
22	62.2	3.5	1290	10	US-09-815-242-7834	Sequence 7834, Ap
23	61.6	3.5	88421	9	US-09-976-059-1	Sequence 1, Appli
24	60.8	3.5	1185	10	US-09-887-576-784	Sequence 784, App
25	60.8	3.5	2787	9	US-09-860-846-40	Sequence 40, Appli
26	60.8	3.5	2787	10	US-09-861-289-40	Sequence 40, Appli
27	60.4	3.4	4689	9	US-09-860-846-34	Sequence 34, Appli
28	60.4	3.4	4689	10	US-09-861-289-34	Sequence 34, Appli
29	60	3.4	7419	10	US-09-815-242-4009	Sequence 4009, Ap
30	59.8	3.4	1815	10	US-09-758-269-13	Sequence 13, Appli
31	59.8	3.4	1839	10	US-09-815-242-4006	Sequence 4006, Ap
32	58.6	3.3	1107	10	US-09-748-033-6	Sequence 6, Appli
33	58.4	3.3	1038	10	US-09-815-242-4048	Sequence 4048, Ap
34	58.2	3.3	2280	10	US-09-815-242-4005	Sequence 4005, Ap
35	57.2	3.3	1325	10	US-09-770-517C-12	Sequence 12, Appli
36	57.2	3.3	1677	10	US-09-770-517C-1	Sequence 1, Appli
37	57.2	3.3	1897	10	US-09-770-517C-25	Sequence 25, Appli
38	57.2	3.3	6854	9	US-09-922-683-7	Sequence 7, Appli
39	56.6	3.2	1533	9	US-09-712-363-81	Sequence 81, Appli
40	55.8	3.2	1374	10	US-09-815-242-7850	Sequence 7850, Ap
41	55.8	3.2	4826	10	US-09-772-304A-1	Sequence 1, Appli
42	55.6	3.2	1203	9	US-09-879-312-1	Sequence 1, Appli
43	55.4	3.2	1224	10	US-09-815-242-7867	Sequence 7867, Ap
44	55.4	3.2	1428	10	US-09-815-242-4000	Sequence 4000, Ap
45	55.2	3.1	798	9	US-09-813-453A-24	Sequence 24, Appli

## ALIGNMENTS

RESULT 1  
US-09-976-059-1  
Sequence 1, Application US/09976059  
Patent No. US20020164747A1  
GENERAL INFORMATION:  
APPLICANT: Farnet, Chris  
APPLICANT: Zazopoulos, Emmanuel  
APPLICANT: Staffa, Alfredo  
TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin  
FILE REFERENCE: 3019-PC1  
CURRENT APPLICATION NUMBER: US/09/976,059  
CURRENT FILING DATE: 2001-10-15  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 88421  
TYPE: DNA  
ORGANISM: Actinoplanes sp.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2077)..(3078)  
OTHER INFORMATION: ORF 1: positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (3118)..(4032)  
OTHER INFORMATION: ORF 2: positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (4038)..(5048)  
OTHER INFORMATION: ORF 3: positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (4038)..(5048)  
OTHER INFORMATION: ORF 3: positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (4038)..(5048)  
OTHER INFORMATION: ORF 3: positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (6665)..(5814)  
OTHER INFORMATION: ORF 4: negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (7703)..(6693)  
OTHER INFORMATION: ORF 5: negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (9464)..(8130)  
OTHER INFORMATION: ORF 6: negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (9691)..(10761)

OTHER INFORMATION: ORF 7; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (12751)..(10829)  
OTHER INFORMATION: ORF 8; negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (13617)..(12802)  
OTHER INFORMATION: ORF 9; negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (15203)..(13614)  
OTHER INFORMATION: ORF 10; negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (15591)..(15863)  
OTHER INFORMATION: ORF 11; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (15880)..(19035)  
OTHER INFORMATION: ORF 12; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (19032)..(39713)  
OTHER INFORMATION: ORF 13; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (39713)..(65800)  
OTHER INFORMATION: ORF 14; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (65826)..(66530)  
OTHER INFORMATION: ORF 15; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (66346)..(67370)  
OTHER INFORMATION: ORF 16; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (67384)..(70059)  
OTHER INFORMATION: ORF 17; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (70099)..(70662)  
OTHER INFORMATION: ORF 18; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (70659)..(71906)  
OTHER INFORMATION: ORF 19; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (73439)..(71964)  
OTHER INFORMATION: ORF 20; negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (74216)..(73563)  
OTHER INFORMATION: ORF 21; negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (73424)..(74213)  
OTHER INFORMATION: ORF 22; negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (75535)..(76464)  
OTHER INFORMATION: ORF 23; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (78110)..(76449)  
OTHER INFORMATION: ORF 24; negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (79864)..(78107)  
OTHER INFORMATION: ORF 25; negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (81624)..(79861)  
OTHER INFORMATION: ORF 26; negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (81909)..(81682)  
OTHER INFORMATION: ORF 27; negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (82346)..(82062)  
OTHER INFORMATION: ORF 28; negative strandedness  
NAME/KEY: misc\_feature  
LOCATION: (82587)..(84446)  
OTHER INFORMATION: ORF 29; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (84461)..(85548)  
OTHER INFORMATION: ORF 30; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (85556)..(86845)  
OTHER INFORMATION: ORF 31; positive strandedness

NAME/KEY: misc\_feature  
LOCATION: (87372)..(86803)  
OTHER INFORMATION: ORF 32; positive strandedness  
NAME/KEY: misc\_feature  
LOCATION: (87494)..(88420)  
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only  
US-09-976-059-1

Query Match 4.9%; Score 86; DB 9; Length 88421;

Best Local Similarity 43.1%; Pred. No. 7,5e-08; Matches 605; Conservative 0; Mismatches 775; Indels 24; Gaps 3;

35 TGTCCATTCGATTCCTCCAGCCCTTGCACCTGCTCTGCGGGGCGACCTCATGACG 94  
Db 46230 TGGCCCGGACGAGGTGCTGCGCAGCGTGTTCACGCTGCGGACGCGGACCGGTGCAAC 46289  
95 GCAGCAACACCCCGGGGGGGGGGGCGGACCTGCGGCTCCGGCGGACCGCATCCGGCGCA 154  
Db 46290 ACATCTTGACCCCGCGGGGGGGGGCGGACCGGCTGCGGCTGCGGACCGGCGGCGC 46349  
155 TCGGCGATCTGTGAGACCGCGCGCGCACACCGGCTGAGCTGTGCGGCGGTGTCG 214  
Db 46350 GGGTGGAGGAGGCGGTCCGCGAAGGGCGCGGTACGCTTGACCTGGCCGGGAGATCC 46409  
215 CGCCCGCTTATGATGATGCGACACCCACGACGACACTTACTGCTAG-----GCGTC 268  
Db 46410 CGCTCGTGGCGTCTGCTGCTGCGCCCGCGGACGACGACGACGACGCTGCTGCTGTC 46469  
269 GCGACATACGCGCCCAAGATCTGCGAGGGGCGTACCAAGGTGTGACGGGCAATTCGGCA 328  
Db 46470 ACCACATGCGCGCGCGCGCGGTGCTGATGCGGCGCTGCGCGCGCGCGCGCGCTT 46529  
329 TCAGCCTGCGCGCGCGTGGCGCGCACCGCCCGCGCGCGCGCGCGCGCGCGCGCGCG 388  
Db 46530 ACCG 46589  
389 GCGGCTTACCGCTTTCAGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 448  
Db 46590 ACTACGCGCTTTCG 46649  
449 CG 508  
Db 46650 CCG 46709  
509 ACTTCACG 568  
Db 46710 CCG 46769  
569 CCATGCG 628  
Db 46770 TCCCG 46829  
629 CCACACCGAGAGATCATTCAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 688  
Db 46830 TCATACCGTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46889  
689 CCACCCACATGCG 748  
Db 46890 TCCGATGCG 46949  
749 TCG 808  
Db 46950 TCTTGTCAACAGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47009  
809 ATTTCG 868  
Db 47010 TGTCTGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47069  
869 TCTCGTGGAGCGCGTATCGTACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 928  
Db 47070 AGAAGCTGTGCGAGAGCTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47123  
929 TGTGCG 988



```
Db 47124 TCAGAGTCATGTCACCTCGACGCGCGCGCGCGAGCGCGCGAGCTGCCGGCCCTGG 47183
Oy 989 ACCTGATGAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1048
Db 47184 CGATGTCGTCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47243
Oy 1049 AGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1108
Db 47244 CCGAGACTTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47303
Oy 1109 TCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1168
Db 47304 CCGAGACTTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47363
Oy 1169 TGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1228
Db 47364 GCGTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47411
Oy 1229 TCGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1288
Db 47412 TCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47471
Oy 1289 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1348
Db 47472 CCGAGATCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47531
Oy 1349 CCGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1408
Db 47532 CCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47591
Oy 1409 TCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47591
Db 47592 TCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47615

RESULT 2
US-09-860-846-7
; Sequence 7, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-7

Query Match 4.7%; Score 81.8; DB 9; Length 1248;
Best Local Similarity 44.2%; Pred. No. 4e-07;
Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;
```

```
Db 335 TCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 394
Oy 212 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 271
Db 395 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 454
Oy 272 ACATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 331
Db 455 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 514
Oy 332 GCGTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 391
Db 515 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 574
Oy 392 GCGTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 451
Db 575 AGGTCCTGAGCTTCA---CGCACAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 631
Oy 452 CCGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 511
Db 632 TCACCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 691
Oy 512 TCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 571
Db 692 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 751
Oy 572 TCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 631
Db 752 TCGGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 811
Oy 632 CCACCGAGAGATCATGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 691
Db 812 CCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 871
Oy 692 CCGACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 751
Db 872 ACGGGCTCAACACACACACACACACACACACACACACACACACACACACACACACAC 931
Oy 752 GCGCGAGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 811
Db 932 ACCCGACCTGTCTATGAGGTCTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 991
Oy 812 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 862
Db 992 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1051
Oy 863 AGGAGTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 922
Db 1052 CCGAGACGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1111
Oy 923 GCGTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 982
Db 1112 ACATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1171
Oy 983 GCGCGAGCTGATGAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015
Db 1172 CCGCGCACCGCGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1204
```

```
US-09-861-289-7
; Sequence 7, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
```

PRIOR FILING DATE: 1998-06-26  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 7  
LENGTH: 1248  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-861-289-7

Query Match 4.7%; Score 81.8; DB 10; Length 1248;  
Best Local Similarity 44.2%; Pred. No: 4e-07;  
Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

```
32 CCATGTCCTCCATTCGATTCGACGCTTCGACCTGCTCTGCGGGGCGGACCCCTCATCG 91
Db 215 CCACTGCAAGGACGACGCGCGGCTCCAGCTCTGCTGCGACGCGCGCTCAGCCGCG 274
Qy 92 AGCGGACGACACCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151
Db 275 AAGGATATATCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 334
Qy 152 CCAATGCGGATCTGTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 211
Db 335 TCACCGCGGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 394
Qy 212 TCGGCGCGCGCTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 271
Db 395 CCGCGCGGCTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 454
Qy 272 ACATGACGCGCGGATCTGCGAGCGCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCG 331
Db 455 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 514
Qy 332 GCTGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 391
Db 515 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 574
Qy 392 GCTGCTACCTGTTGAGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 451
Db 575 AGGCTTACGCTTCA---CGCACCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 631
Qy 452 CCGTCAAGCGCGCGCTGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 511
Db 632 TCACCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 691
Qy 512 TCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 571
Db 692 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 751
Qy 572 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 631
Db 752 TGGGCTTACCTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 811
Qy 632 CCACGAGAGATCATCGAGGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 691
Db 812 CCGCTTACCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 871
Qy 692 CCCACATCGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 751
Db 872 ACGGCTCAACACACACAGTACGTGATGATGATGATGATGATGATGATGATGATGAT 931
Qy 752 GCGCGGAGCTGAGCGTGGCGGTGATGATGATGATGATGATGATGATGATGATGATGAT 811
Db 932 ACCGCGACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 991
Qy 812 TCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 862
Db 992 CGCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1051
Qy 863 AGGAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 922
Db 1052 CCGAAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1111
```

```
Qy 923 GCGTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 982
Db 1112 ACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1171
Qy 983 GCGCGGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1015
Db 1172 CGCGCGACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1204
```

## RESULT 4

US-09-860-846-3/c  
Sequence No. Application US/09860846  
Patent No. US20020164742A1

## GENERAL INFORMATION:

APPLICANT: Sherman, D.H.

APPLICANT: Liu, H.

APPLICANT: Xue, Y.

APPLICANT: Zhao, L.

TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600.438051

CURRENT APPLICATION NUMBER: US/09/860,846

PRIOR FILING DATE: 2001-05-18

PRIOR APPLICATION NUMBER: 09/105,537

PRIOR FILING DATE: 1998-06-26

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 3

LENGTH: 13613

TYPE: DNA

ORGANISM: Streptomyces venezuelae

US-09-860-846-3

Query Match 4.7%; Score 81.8; DB 9; Length 13613;  
Best Local Similarity 44.2%; Pred. No: 4.3e-07;  
Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

```
Qy 32 CCAATGTCCTCCATTCGATTCGACGCTTCGACCTGCTCTGCGGGGCGGACCCCTCATCG 91
Db 12669 CCACTGCAAGGACGACGCGCGGCTCCAGCTCTGCTGCGACGCGCGCGCTCAGCCGCG 12610
Qy 92 AGCGGACGACACCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151
Db 12609 AAGTATATATCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 12550
Qy 152 CCAATGCGGATCTGTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 211
Db 12549 TCACCGCGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12490
Qy 212 TCGCGCGCGCGCTTCAATGACTGCGACACCGAGAGACACTACTGCTCAAGCGTGGCG 271
Db 12489 CCGCGCGGTCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12430
Qy 272 ACATGACGCGCGGATCTGCGAGCGCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 331
Db 12429 GCGCGCGGACCAAGCTGCGGAGGTCGCGAGACGACGCGCGCGCGCGCGCGCGCGCG 12370
Qy 332 GCTGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 391
Db 12369 CCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12310
Qy 392 GCTTACCTGTTGAGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 451
Db 12309 AGGCTTACGCTTCA---CGCACCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12253
Qy 452 CCGTCAAGCGCGCGCTGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 511
Db 12252 TCACCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12193
Qy 512 TCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 571
Db 12192 TCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12133
Qy 572 TGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 631
```

```
Db 12132 TGGGCTACCTCCCTGAGCGGCTTCCCGAGTGCATGACACCGGACCGGACACACG 12073
QY 632 CCACCGAAGATATCATGAGTGTGCGGCGCGTGAAGCGCCATGCGGCGATCTACGCCA 691
Db 12072 CCGCTTACCGGAGACACCTCCCGGAGCTCCCGGCTGCTGCGGCGACACGACCGCC 12013
QY 692 CCAATGCTCGGAGGAGGAGGACATCGGCGCGCGCTGAGGAAACCTTCCGCAATG 751
Db 12012 AGGCGCTAACACACACAGTACGTGATGATGAGATGACAGACGCGACCGCATCC 11953
QY 752 GCGCGGAGCTGAGATGCGCGGTGATCTGACACCAAGATGAGCGGACCGCAAT 811
Db 11952 ACCGAGACCTGTGATGAGAGTCTGAGAGCGGAGGCTGACACCGCGCTACTTCT 11893
QY 812 TCGGCGCTCGCGGAGACGCTGCCGCTGATGAG-----GCGCGCATGCGCGCC 862
Db 11892 CGCGGCGCTGCGGAGCTGAGAGCGCTACCGGCGGACCGCGCGCGCTGCGCGACA 11833
QY 863 AGAGCTCTGCTGAGACGCTATCCCTACGTGAGCGCGCTCCACCATGCTCAAGAGGAC 922
Db 11832 CCGAAGCGCTCGCGCGCGCGGCTGCTGCTGCTGCGACCGGACCGCATCGCGGACG 11773
QY 923 GCGTGTGCTGCGCGGAGACCATCATCATCTGATGAGCGCTTCCCGGAGTGAAGCG 982
Db 11772 ACATCGCGCGGAGTGGCGGACCTGCTGCTGCTGCGGAGCGCGCGCGGAGTGAAGCG 11713
QY 983 GCGCGACCTGATGAGTCCGCGCGGCGGCGG 1015
Db 11712 CGCGCACCGGAGACGCGCGCGCGCGCTCG 11680
```

## RESULT 5

```
US-09-861-289-3/c
; Sequence 3, Application US/09861289
; Patent No. US2002011087A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding myxymycin and plixymycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-3
```

```
Query Match 4.7% Score 81.8; DB 10; Length 13613;
Best Local Similarity 44.2%; Pred. No. 4.3e-07;
Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;
```

```
QY 32 CCATGTCCTCAATCCGATTCAGAGCCCTTGAAGCTGCTGCTGCGGCGGCGACCTCATG 91
Db 12669 CCACCTGCAACGCGACGCGCGGCTCCAGCTCTGCGGACGCGCGCGCTCAACGCGCG 12610
QY 92 ACGGAGCAACACCGCGGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 151
Db 12609 AAGTATATATCGCTGATGATGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12550
QY 152 CCATGCGGATCTGTGCGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 211
Db 12549 TCACCGCGGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12490
QY 212 TCGCGCGCGGCTTCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 271
```

```
Db 12489 CCGCGCGGCTGACACCGCGGACCTCGGCGGCTGCGGCGGCTGACACCTTGGGCGCGCCCT 12430
QY 272 ACATGAGCGCCAGATCTCCAGGCGCTACACAGGCTGTGATCAGGCGCATTTGGCGCATCA 331
Db 12429 GCGCGCGGCGGACGCTGCGGAGAGTGTGCGGAGGACGAGGCGCTGCGGCTGATCGAGC 12370
QY 332 GCGTGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 391
Db 12369 CCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 12310
QY 392 GCTTTCACCTTTCGAGCGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTC 451
Db 12309 AGTCTTTCAGCTTTC-----GCGGACCAAGCGCGGCGGCGGCGGCGGCGGCGGCGG 12253
QY 452 CCGTCAACCGCGGCTGATGAGTGGCGGCTTTCAGCGCTGCGGCGGCGGCGGCGGCGGCGG 511
Db 12252 TCACCGAGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12193
QY 512 TCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 571
Db 12192 TCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12133
QY 572 TGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 631
Db 12132 TGGCGCTACCTCTCGGAGCGGCTTTCGAGGCTGATGAGCGGCGGCGGCGGCGGCGG 12073
QY 632 CCACCGAAGATATGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 691
Db 12072 CCGCTTACCGGAGACACCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12013
QY 692 CCGACATGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 751
Db 12012 ACAGGCTTCAACACACAGTACGTGATGATGAGATGAGAGGCGGCGGCGGCGGCGG 11953
QY 752 GCGCGAGCTGAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 811
Db 11952 ACCGAGACCTGTGATGAGAGTCTGAGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGG 11893
QY 812 TCGGCGCTCGCGGAGACGCTGCCGCTGATGAG-----GCGCGCATGCGCGCC 862
Db 11892 CGCGGCGCTGCGGAGCTGAGAGCGGCTACCGGCGGCGGCGGCGGCGGCGGCGGCGG 11833
QY 863 AGAGCTCTGCTGAGACGCTATCCCTACGTGAGCGCGCTCCACCATGCTCAAGAGGAC 922
Db 11832 CCGAAGCTTTCGCGCGCGGCGGCGGCTGCTGCTGCTGCGACCGGACCGCATCGGCGG 11773
QY 923 GCGTGTGCTGCGCGGAGACCATCATCATCTGATGAGCGCTTCCCGGAGTGAAGCG 982
Db 11772 ACATCGCGCGGAGTGGCGGACCTGCTGCTGCTGCTGCGGAGCGGCGGCGGCGGCGG 11713
QY 983 GCGCGACCTGATGAGTCCGCGCGGCGGCGG 1015
Db 11712 CGCGCACCGGAGACGCGCGCGCGCGCTCG 11680
```

## RESULT 6

```
US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding myxymycin and plixymycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
```

LENGTH: 15872  
 TYPE: DNA  
 ORGANISM: Streptomyces venezuelae  
 US-09-861-289-1

Query Match 4.2%; Score 73.8; DB 9; Length 15872;  
 Best Local Similarity 43.2%; Pred. No. 1.3e-05;  
 Matches 586; Conservative 0; Mismatches 747; Indels 24; Gaps 4;

OY 175 GCGCGCACACCGGGTGCACGTGTGGGCTGTGTCGGCCCGCTTCATCAGCTG 234  
 DB 4799 GGGAGAGAACTGGCGCGCGGACACCCCTGTTCGCGCCCTCGACACCGTACG 4858  
 OY 235 CACACCCAGCAGCAACTGTCTGCTGAGCGTGGAGATGAGCGCCAGATCTGCGAG 294  
 DB 4859 GCGCTGACCGCTACCTCTGACCGCGCGCTGCGGAGATCTCTCGCCCGGGAGAGACTG 4218  
 OY 295 GCGCTGACCAAGGTGTGACAGGGCAATGGCGCATGAGCTGCGCGCTGGCGGACGCG 354  
 DB 4919 GACTTCACCGCTGACACCGACCGCGCTTCTGCGCTTTCAGAGTGGCGTGTTCGCGCTC 4978  
 OY 355 AACCCGCGCGCGCGCTGACCTGTGTGAGAGAGAGGCGCTTACCTTTCAGAGCTTC 414  
 DB 4979 CTGGAACACACGCGCTGTGCGGACCTGTGACCTGTCTACCGCTGCGGAGATCGCC 5038  
 OY 415 GCGGACTACCTGAGCGGCTGTGGGCGACGCGCGCGCTCAACGCGCGCTGTATGCTG 474  
 DB 5039 GCGCGGACGCTGCGCGGTCTCTCTCCCTGAGACGCGCGCAGCTCTGTACCGCGCGC 5098  
 OY 475 GCGCATTCACGCTGCGCGCGCGGTATCGCGACTTTCAGCGCGCGCGCACCGACGAG 534  
 DB 5099 GCGCGGCTCATGCTGCGCGCGCGGAGGCGCGCGCTGATTCGCGCTGACGCGCGCGAG 5158  
 OY 535 GAATTCGCGCGCATGCGGAGACCTGCGGAGAGAGCATGCGCGCGCGCATGCGGCTT 594  
 DB 5159 GCGAGGCTCTGCTGAGTCCCTGAGAGGCTGAGAGGCGCAGGCTGCGCGCGCTGAC 5218  
 OY 595 TCGACCGCGCGCTTCTACCGCGCGCGCGCGCGCGCGCACCAAGAGATCATGAGAGTG 654  
 DB 5219 GGACCGACCGCGCGGTGCTGCTCGCGGAGCGGAGCGCGCGCGAGAGTCCGCGCTA 5278  
 OY 655 TCGCGCGCGCTGAGCGCGCGCTGCGCGCTACCGCGCACCATGCGCGGAGAGAGGAG 714  
 DB 5279 TGGCGGAGACCGCGCGCGCGCGCGCGAGGCTGCGCTACCGCGCTTTCACCTCCCG 5338  
 OY 715 CACATCGTGGCGCGCTGAGAGAACTTCGCTGCGCGCGCGCGCTGAGAGCTGCGGCTG 774  
 DB 5339 CACATGAGAGAGCTCTGAGAGAGTCTCTCGGCTCGCGCGCGCGCTGAGAGAG 5398  
 OY 775 GTGATCTCGACCAAGTCTATGGCGACGCCAATTCGCGCGCTGCGCGGAGAGCGCTG 834  
 DB 5399 CCGCGGATCCCGCTGCTCTCCAGGCTACCGGCGCGCTGCTCAGCTCCGCGGA----- 5451  
 OY 835 CCGCTGATCGAGCGCGCGCTGCGCGCGCGCGCGCGCGCTGCGCTGAGAGCGGTATCTACG 894  
 DB 5452 --GCTCACCTGCGCGCGCTGCTGAGTCCGAGATCCGCGCGCGCGCTTCTCTGAGC 5509  
 OY 895 GCGCGCTCACCATCTGCTCAACGAGACGCGCGCTGCTGCGCGCGCGCGCGCATCATCACC 954  
 DB 5510 GCGCTCGCGACCTCTGCG 5569  
 OY 955 TGGTGCAGAGCGCTTCCCGAATGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1014  
 DB 5570 GTCTCTACGCGGCTGCGGAGAGGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5629  
 OY 1015 GCGCAATCAAGTACGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074  
 DB 5630 GACGTCAAGGCTGCTGCG 5689  
 OY 1075 ATGACGAGACCGAGCTGACGCGCATCTGCGGTTCGCGCGCGCGCGCGCGCGCGCGCG 1134  
 DB 5690 GGTCTCGGAGCG 5749

OY 1135 GCGCTGCG 1194  
 DB 5750 GCG 5809  
 OY 1195 GCG 1248  
 DB 5810 GAGGCG 5869  
 OY 1249 ATGACG 1302  
 DB 5870 ACCG 5929  
 OY 1303 GCGCGGCTACT--TCG 1359  
 DB 5930 GACGACCG 5989  
 OY 1360 ACCTTGAGACCGCTTACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1419  
 DB 5990 TTTCTGAGGCTGCG 6049  
 OY 1420 CCGGCTGCGGAGAGGCGGTTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1479  
 DB 6050 ACCCTGAGAGCG 6109  
 OY 1480 GCGCGCTGAGCG 1516  
 DB 6110 AGCGGAGAGCGGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6146

RESULT 7  
 US-09-861-289-1  
 ; Sequence 1, Application US/09861289  
 ; Patent No. US20020110897A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600.438US1  
 ; CURRENT APPLICATION NUMBER: US/09/861,289  
 ; PRIORITY FILING DATE: 2001-05-18  
 ; PRIOR APPLICATION NUMBER: 09/105,537  
 ; PRIOR FILING DATE: 1998-06-26  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 15872  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 US-09-861-289-1

Query Match 4.2%; Score 73.8; DB 10; Length 15872;  
 Best Local Similarity 43.2%; Pred. No. 1.3e-05;  
 Matches 586; Conservative 0; Mismatches 747; Indels 24; Gaps 4;

OY 175 GCGCGCACACCGGGTGCACGTGTGGGCTGTGTCGGCCCGCTTCATCAGACTG 234  
 DB 4799 GGGAGAGAACTGGCGCGCGGACACCCCTGTTCGCGCCCTCGACACCGTACG 4858  
 OY 235 CACACCCAGCAGCAACTGTCTGCTGAGCGTGGAGATGAGCGCCAGATCTGCGAG 294  
 DB 4859 GCGCTGACCGCTACCTCTGACCGCGCGCTGCGGAGATCTCTCGCCCGGGAGAGACTG 4918  
 OY 295 GCGCTGACCAAGGTGTGACAGGGCAATGGCGCATGAGCTGCGCGCGCTGGCGGACGCG 354  
 DB 4919 GACTTCACCGCTGACACCGACCGCGCTTCTGCGCTTTCAGAGTGGCGCTTTCGCGCTC 4978  
 OY 355 AACCCGCGCGCGCGCTGACCTGTGTGAGAGAGGCGCTTACCTTTCAGAGCTTC 414  
 DB 4979 CTGGAACACACGCGCTGTGCGGACCTGTGACCTGTCTACCGCGCGCGCGCGAGATCGCG 5038  
 OY 415 GCGGACTACCTGAGCGGCTTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTATGCTG 474







QY 796 ATGGCCACCCCAATTTGCGCGCGCGAGACGTCCGCTGATGAGGCGGCATG 855  
 Db 1093 TGGGGCCCTCGATCATCCCGTGGCCCGGCGGTGTGCTGATGAGCACTTCCGC 1152  
 QY 856 GCGGCGCA 863  
 Db 1153 GTGATGCA 1160

# RESULT 10 US-09-860-846-30

; Sequence 30, Application US/09860846  
 ; Patent No. US20020164742A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600,438051  
 ; CURRENT APPLICATION NUMBER: US/09/860,846  
 ; PRIOR FILING DATE: 2001-05-18  
 ; PRIOR APPLICATION NUMBER: 09/105,537  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FASTSEQ for Windows Version 3.0  
 ; SEQ ID NO 30  
 ; LENGTH: 13842  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 ; US-09-860-846-30

Query Match 4.1%; Score 72.2; DB 9; Length 13842;  
 Best Local Similarity 44.8%; Pred. No. 2.5e-05;  
 Matches 413; Conservative 0; Mismatches 493; Indels 15; Gaps 3;

QY 76 GCGGCAACCTCATGACGCGACACACCCCGGCGCGCGCGACCTGGCGTGGC 135  
 Db 463 GTCCACCGGTTGACGCGCGCTTCTGCGCATTCGCCCCCGAGAGATGAC 522  
 QY 136 GCGCACCGCATCGCGCGCATGCGGATCTGTGAGACCGCGCGACACCGGCTGAC 195  
 Db 523 CCGGAGAGCGCGTCTGCGCTGAGCTGGGAGGCGCTGAGAGCGCGCGATGAC 582  
 QY 196 GTGCGGCGCTGTGTGCGCGC---CCGGCTTCATGACCTGCGACACCGACGACGAC 252  
 Db 583 CCGTCTCTGCTACCGCGACCGACCGCGCGCTTCTGCGCGCTGCGGCTATCTGGAGACATGAC 642  
 QY 253 TACCTGCTCAGGCGCTGCGCATGACGCCCAAGATCTGCGAGGCGTCCACCGATGCTG 312  
 Db 643 GCCACCTTGAAGACCGCGAGGCGCGCGCGCGCATCACCGCGACCGTACCGGCGCTC 702  
 QY 313 ACGGCGAATTCGCGCATGACCTGCGCGCGCGCGCGACCGCAACCGCGCGCGCGCTG 372  
 Db 703 CACCGCGCATCATGCGGAGACCGACTCTGTAACAGCTCGGCGCTCGCGCGCGCGCATG 762  
 QY 373 GACCTGTGAGCGAAGCGGCTTACCGTTTCAGAGCGCTTCGCGCATCTGAGCGCG 432  
 Db 763 GTCTGTCACTCCGCGCACTCTCTGCTCTGCTCGCTTCACCTCGCTGCGAGAGCTG 822  
 QY 433 TTGCGGCGCAACGCGCGCGCGCTCAACGCGCGCTGTATGTTGGGCGCATTAAGCTGGC 492  
 Db 823 CCGGCGCGCGAGTCCGAGCTGCGCGCGCGCGCGCGCGCTGCTGCTAACCTGTGCGGAC 882  
 QY 493 GCGCGGCTCATGCGCGACTTCAGCGCGCGCGCGCGACCGAGAAATCGCGCGCGCGCG 552  
 Db 883 AGCATCATTCGCGCGGAGCAAGTTGCGCGCGCTTCCCGCGCGCGCGCGCTACACCTTC 942  
 QY 553 GACCTGCGCGAGAGCAATGCGCGCGCGCGCGCGCGCGCATGTTGAGCGCGCGCGCTTAC 612  
 Db 943 GACCGCGCGCGCAAGCTGTAAGTACGCGCGCGCGCGCGCGGTTTCTGCTTCAAGCGC 1002  
 QY 613 CCG 672

Db 1003 CTCTCCCGCGCGTGGCGGAGCGGACCCGCTGCTCGCGTATCGGCGCGCGCGCTC 1062  
 QY 673 CATGCGCGCATCTACCGCGCACCATATGCGCGAGAGAGGACATGTCGCGCGCTG 732  
 Db 1063 AACAAAGCG 1113  
 QY 733 GAGGAACTTCCCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 792  
 Db 1114 GAGCGCGTCTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1173  
 QY 793 GTCATGCG 852  
 Db 1174 GTGAGCTGACCG 1233  
 QY 853 ATGCG 912  
 Db 1234 GCGCGCGCTCG 1293  
 QY 913 AAGCAGACCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 972  
 Db 1294 AACATCG 1350  
 QY 973 GACCTGAGCG 993  
 Db 1351 CCGGTCG 1371

## RESULT 11 US-09-861-289-30

; Sequence 30, Application US/09861289  
 ; Patent No. US20020110897A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600,438051  
 ; CURRENT APPLICATION NUMBER: US/09/861,289  
 ; PRIOR FILING DATE: 2001-05-18  
 ; PRIOR APPLICATION NUMBER: 09/105,537  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FASTSEQ for Windows Version 3.0  
 ; SEQ ID NO 30  
 ; LENGTH: 13842  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 ; US-09-861-289-30

Query Match 4.1%; Score 72.2; DB 10; Length 13842;  
 Best Local Similarity 44.8%; Pred. No. 2.5e-05;  
 Matches 413; Conservative 0; Mismatches 493; Indels 15; Gaps 3;

QY 76 GCGGCAACCTCATGACGCGACACACCCCGGCGCGCGCGCGCGACCTGGCGTGGC 135  
 Db 463 GTCCACCGGTTGACGCGCGCTTCTGCGCATTCGCCCCCGAGAGATGAC 522  
 QY 136 GCGCACCGCATCGCGCGCATGCGGATCTGTGAGACCGCGCGACACCGGCTGAC 195  
 Db 523 CCGGAGAGCGCGTCTGCGCTGAGCTGGGAGGCGCTGAGAGCGCGCGCGATGAC 582  
 QY 196 GTGCGGCGCTGTGTGCGCGC---CCGGCTTCATGACCTGCGACACCGAGAGAC 252  
 Db 583 CCGTCTCTGCTACCGCGACCGCGCGCGCGCGCGCGCTTCTGCGCGCGCGCGCGCG 642  
 QY 253 TACCTGCTCAGGCGCTGCGCATGACGCCCAAGATCTGCGAGGCGCTTACCGAGTGTG 312  
 Db 643 GCCACCTTGAAGACCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 702  
 QY 313 ACGGCGAATTCGCGCATGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372

Db 703 CACCGGCGATCATCGGAAACGACTCTGTACAGCTCGGGCTCCGGGCCACGATG 762  
OY 373 GACCTGTGAGCAAGAGGGGCTTTACCGGTTTCAGCGCTTCGGCGACTACGTGAGCGC 432  
Db 763 GTCTGTGACTTCGGGCGACTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 822  
OY 433 TTGCGGCGCAAGCGGCGGCGCTGACAGCGGCTGTATGTGTGGCGCTTCAAGCGCTGCG 492  
Db 823 CGGCGGCGGCGAGTCCGAGCTGCGGCTGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTG 882  
OY 493 GCGCGGCTGATCGGCGACTTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 552  
Db 883 AGCATCTATCGGCGGCGAGCAAGTTCGCGGCGCTCTGCGGCGGCGGCGGCGGCGGCGG 942  
OY 553 GACCTGTGAGCAAGAGGGGCTTCAGCGGCTTCAGCGGCTTCAGCGGCTTCAGCGGCTTC 612  
Db 943 GACGCGCGCGCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1002  
OY 613 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 672  
Db 1003 CTCTCCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1062  
OY 673 CATGCGGCGATCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 732  
Db 1063 AACAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1113  
OY 733 GAGGAAACCTTCCGCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 792  
Db 1114 GAGCGCGGCTGCTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1173  
OY 793 GTCTATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 852  
Db 1174 GTGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1233  
OY 853 ATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 912  
Db 1234 GCGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1293  
OY 913 AAGCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 972  
Db 1294 AACATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1350  
OY 973 GAACGTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 993  
Db 1351 GCGGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1371

## RESULT 12

US-09-860-846-5  
; Sequence 5, Application US/09860846  
; Patent No. US20020164742A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D. H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438051  
; CURRENT APPLICATION NUMBER: US/09/860.846  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105.537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 36778  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-860-846-5

Query Match 4.1%; Score 72.2; DB 9; Length 36778;  
Best Local Similarity 44.8%; Pred. No. 2.5e-05;  
Matches 413; Conservative 0; Mismatches 493; Indels 15; Gaps 3;

OY 76 GCGGCGACCCCTCATCGACGCGGCAACACCCCGGGCGGCGGCGGCGGCGGCGGCGGCGG 135  
Db 2204 GTGACCGGCTGCGAGCGCGCGCTTCTGCGGATCTGCGCCCGGAGGCGCGGAGATGAGC 2263  
OY 136 GCGGACCGCATGCGCGCATTCGCGGATCTGTGAGACGCGCGCGCGGCGGCGGCGGCGG 195  
Db 2264 CCGGACCGACCGGCTGCGCGCTGCGGAGCTGCGGAGCGGCTGCGGAGCGGCGGCGGAGTCAG 2323  
OY 196 GTGCGGGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 252  
Db 2324 CCGTCTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2383  
OY 253 TACCTGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 312  
Db 2384 GCGACCGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2443  
OY 313 ACAGGCGATTTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 372  
Db 2444 CACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2503  
OY 373 GACCTGTGAGCAAGAGGGGCTTCAGCGGCTTCAGCGGCTTCAGCGGCTTCAGCGGCTTC 432  
Db 2504 GTGCTGACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2563  
OY 433 TTGCGGCGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 492  
Db 2564 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2623  
OY 493 GCGCGGCTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 552  
Db 2624 AGCATCTATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2683  
OY 553 GACCTGTGAGCAAGAGGGGCTTCAGCGGCTTCAGCGGCTTCAGCGGCTTCAGCGGCTTC 612  
Db 2684 GAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2743  
OY 613 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 672  
Db 2744 CTCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2803  
OY 673 CATGCGGCGATCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 732  
Db 2804 AACAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2854  
OY 733 GAGGAAACCTTCCGCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 792  
Db 2855 GAGCGCGTGTGCTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2914  
OY 793 GTCTATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 852  
Db 2915 GTGAGCTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2974  
OY 853 ATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 912  
Db 2975 GCGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3034  
OY 913 AAGCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 972  
Db 3035 AACATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3091  
OY 973 GAACGTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 993  
Db 3092 GCGGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3112

## RESULT 13

US-09-861-289-5  
; Sequence 5, Application US/09861289  
; Patent No. US20020110897A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D. H.  
; APPLICANT: Liu, H.

APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
FILE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600,438US1  
CURRENT APPLICATION NUMBER: US/09/861,289  
PRIOR FILING DATE: 2001-05-18  
PRIOR APPLICATION NUMBER: 09/105,537  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 5  
LENGTH: 36778  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-861-289-5

Query Match 4.1%; Score 72.2; DB 10; Length 36778;  
Best Local Similarity 44.8%; Pred. No. 2.5e-05;  
Matches 413; Conservative 0; Mismatches 493; Indels 15; Gaps 3;

Qy 76 GCGGCACTTCATGACGCGACGACACACCCGGGGCGCGCGACCTGGGCGTGGC 135  
Db 2204 GTGCGCCGTTGACGCGCGCTTCTTCCGACATCTGCGCCCGAGCGCGGAGATGAC 2263  
Qy 136 GCGGACCGGATGCGCGGATGCGGATCTGTCGACGCGCGCGCGCGACACCGGGTGCAC 195  
Db 2264 CCGGACGACGCGGCTGCGCGGCTGAGCTGGGCTGGGAGCGCTGGAGCGCGGGATGAC 2323  
Qy 196 GTGCGGCGCTGGTGGTGGCGG---CGGCTTCATGACGCGACACCGACGACGACGAC 252  
Db 2324 CGGCTCGCTGACGCGGACCGCGCGCGCGCTTCCGCGGGGCGCATGTGGAGACGACAC 2383  
Qy 253 TACGCTGACGCGCTGCGGACATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 312  
Db 2384 GCGACCTGACGCTC 2443  
Qy 313 ACGGCAATTCGCGGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 372  
Db 2444 CACGCGCGGATGACGCTG 2503  
Qy 373 GACCTGCTGACGCTG 432  
Db 2504 GTGCTGACGCTG 2563  
Qy 433 TTGCGGCTG 492  
Db 2564 CCGGCGCGGATGCTG 2623  
Qy 493 GCGGCGGATGCTG 552  
Db 2624 AGCATATTCGCTG 2683  
Qy 553 GACCTGCTG 612  
Db 2684 GACGCTG 2743  
Qy 613 CCGCTG 672  
Db 2744 CTCTCCCGCTG 2803  
Qy 673 CATGCTG 732  
Db 2804 AACACGCTG 2854  
Qy 733 GAGGAACTTCGCTG 792  
Db 2855 GAGGCTG 2914  
Qy 793 GTCATGCTG 852  
Db 2915 GTGAGCTGCTG 2974  
Qy 853 ATGGCTG 912

Db 2975 GCGGCTG 3034  
Qy 913 AACGAGCTG 972  
Db 3035 AACGAGCTG 3091  
Qy 973 GACTGAGCTG 993  
Db 3092 GCGGCTG 3112

RESULT 14  
US-09-860-846-32  
Sequence 32, Application US/09860846  
Patent No. US20020164742A1  
GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
FILE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600,438US1  
CURRENT APPLICATION NUMBER: US/09/860,846  
PRIOR FILING DATE: 2001-05-18  
PRIOR APPLICATION NUMBER: 09/105,537  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 32  
LENGTH: 11220  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-860-846-32

Query Match 4.0%; Score 69.6; DB 9; Length 11220;  
Best Local Similarity 45.7%; Pred. No. 7.4e-05;  
Matches 332; Conservative 0; Mismatches 379; Indels 15; Gaps 2;

Qy 441 CACGCTG 500  
Db 10194 CACCTGACCGCTG 10253  
Qy 501 CATCCGCTG 560  
Db 10254 CTTCGCTG 10313  
Qy 561 CGAGGAGCTG 620  
Db 10314 CTGCTG 10373  
Qy 621 CGCTG 680  
Db 10374 GACGAGCTG 10433  
Qy 681 CATCTAGCTG 740  
Db 10434 GCGGCTG 10481  
Qy 741 CTTCGCTG 800  
Db 10482 CATCGCTG 10541  
Qy 801 CACGCTG 860  
Db 10542 CACGCTG 10598  
Qy 861 CAGGAGCTG 920  
Db 10599 GCGGCTG 10658  
Qy 921 CCGGCTG 980



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2003, 06:29:34 ; Search time 74 Seconds  
(without alignments)  
7285.645 Million cell updates/sec

Title: US-10-009-782-1

Sequence: 1 gaattccattgacgcgga.....ccctgagctacgagaagctt 1758

Scoring table: IDENTITY-NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents.NA.\*  
2: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCrus.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/Backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.2	5.4	8438	1	US-07-945-283-1
2	81.8	4.7	1248	4	US-09-105-537-7
3	81.8	4.7	13613	4	US-09-105-537-3
4	77.4	4.4	23673	4	US-09-773-816-1
5	76.2	4.3	5970	3	US-09-320-878-21
6	75.6	4.3	43280	2	US-08-804-227C-1
7	75	4.3	1320	2	US-08-461-775-8
8	75	4.3	1320	2	US-08-461-775-8
9	75	4.3	1620	2	US-08-461-775-10
10	75	4.3	1620	2	US-09-031-606-11
11	74.4	4.2	12001	1	US-08-458-568A-11
12	73.8	4.2	15872	4	US-09-105-537-1
13	73.4	4.2	2167	2	US-08-461-775-9
14	73.4	4.2	2167	2	US-09-031-606-9
15	73.4	4.2	2668	2	US-08-461-775-11
16	73.4	4.2	2668	2	US-09-031-606-11
17	73.4	4.2	4257	2	US-08-690-473-1
18	73.4	4.2	4257	4	US-09-259-821A-1
19	73.4	4.2	4257	4	US-08-843-659-1
20	72.8	4.1	11604	4	US-09-385-028-13
21	72.8	4.1	15079	4	US-09-385-028-1
22	72.2	4.1	13842	4	US-09-105-537-30
23	72.2	4.1	36778	4	US-09-105-537-5
24	72.2	4.1	38506	3	US-09-320-878-19
25	71.2	4.1	1833	2	US-08-403-852D-6
26	71.2	4.1	1833	3	US-08-510-646B-6
27	71.2	4.1	1833	4	US-09-231-818-6

28	71.2	4.1	2634	3	US-08-941-936-1	Sequence 1, Appl1
29	70	4.0	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
30	69.6	4.0	11220	4	US-09-105-537-32	Sequence 32, Appl1
31	68.8	3.9	5392	2	US-08-403-852D-1	Sequence 1, Appl1
32	68.8	3.9	5392	3	US-08-510-646B-1	Sequence 1, Appl1
33	68.8	3.9	5392	4	US-09-231-818-1	Sequence 1, Appl1
34	68.6	3.9	44377	2	US-08-604-227C-7	Sequence 7, Appl1
35	68.6	3.9	44377	2	US-08-804-198-1	Sequence 1, Appl1
36	68.2	3.9	2220	4	US-08-765-907A-14	Sequence 14, Appl1
37	68.2	3.9	4496	4	US-08-765-907A-6	Sequence 6, Appl1
38	68	3.9	2064	1	US-08-343-428-1	Sequence 1, Appl1
39	68	3.9	12588	2	US-08-387-942C-1	Sequence 1, Appl1
40	67.8	3.8	1014	4	US-09-105-537-13	Sequence 13, Appl1
41	67.4	3.8	30001	1	US-08-125-468-1	Sequence 1, Appl1
42	67.4	3.8	30001	2	US-08-474-933-1	Sequence 1, Appl1
43	67.2	3.8	1208	2	US-08-403-852D-4	Sequence 4, Appl1
44	67.2	3.8	1208	3	US-08-510-646B-4	Sequence 4, Appl1
45	67.2	3.8	1208	4	US-09-231-818-4	Sequence 4, Appl1

#### ALIGNMENTS

RESULT 1  
US-07-945-283-1/C  
Sequence 1, Application US/07945283  
Patent No. 5352596  
GENERAL INFORMATION:  
APPLICANT: Cheung, Andrew K.  
APPLICANT: Wesley, Ronald D.  
TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants  
TITLE OF INVENTION: Involving The Ep0 and Lf1 Genes  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis P. Ribando  
STREET: 1815 No. 5352596th University Street  
CITY: Peoria  
STATE: IL  
COUNTRY: USA  
ZIP: 61604  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/945,283  
FILING DATE: 19920911  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ribando, Curtis P.  
REGISTRATION NUMBER: 27976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 309-685-4011 ext. 513  
TELEFAX: 309-685-4128  
INFORMATION FOR SEQ ID NO.: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8438 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Pseudorabies virus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 622..6495  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1099, "g")  
FEATURE:

```

NAME/KEY: variation
LOCATION: replace(1267, "t")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1381, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1566, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(7010, "g")
US-07-945-283-1

```

Query Match	5.4%	Score 94.2;	DB 1;	Length 8438;
Best Local Similarity	44.5%	Pred. No. 8.8e-10;		
Matches 608;	Conservative	0;	Mismatches 738;	Indels 21;
				Gaps 5;

OY	31	TCATATGCCAATCCGATATCCAGACCCCTTCAGACCTGCTCTCGCGGGGAGCAACCCATC	90
Db	5604	TCCTGCTCTCTCTCCGAGGGAGAAAGACGAAAGATCTCGGCCCGGCGCCCACTCGCC	5545
OY	91	GACGCGACGACACCCCGGGGGGCGCGCCGACCTGTGGCGTGCAGCGACCGCATCGCC	150
Db	5544	CGCGCCCGGGCCGCAACCGTCTCCGCGGGGGCCGCGCGCCCGCGGCGCTTCGCGCTCC	5485
OY	151	GCCATTCGGGATCTGTCCGACCGCCCGCGACACCCGGGTGAGCTGTTCGGGCTGGTG	210
Db	5484	TCGCCCTCCGACCACTCTCTCTCTCGCGGGGCTCTCCCGGCCCGCCCGCGCGCGCG	5428
OY	211	GTGCGGCGCGGGCTTCATCTGACTTCGACACCCACGACGACGACATCTGCTGACGCTCGC	270
Db	5427	GCCCGGCGCCCGCGCGAGAAAGCGCGCTCTACCAACACACTCTCTGCTACGCGCGAC	5368
OY	271	GACATGACGCGCCCAAGATCTCGAGAGGGCGTTCACCAACGCTGATCAGGGACATTCGGCATC	330
Db	5367	GGGCGCCCGCGACGACGACGCGCGCTCTACCCCGCTCGGGGAGCGCTGGCCCGGCTCC	5308
OY	331	AGCCTGGCGCGCGCTGGCGGAGCGCAACCGCGCCCGCCCTCGACCTGCTGGACGAGCAAGCG	390
Db	5307	GACCTCTCGCGCGAGCGCGCGCTCTACCGCTGACGGCGCGCGGGGAGCTCCGCGAGGGGCTC	5248
OY	391	GGCTTTACCGTTTCGAGGCGTTTCGCGCACTTACTGTGACCGCTTGGCGGCGACGCGCGCG	450
Db	5247	TGGGACGAGAGAGACGCTCCGCCACGCGCGGGCGCCGCTACCGCGCGCGCGCGCGCGCTG	5188
OY	451	GCCGTCACGCGCGCTGTATGTGGGCAATTCAACGCTGCGCGCGCGGCTGATCGCGGAC	510
Db	5187	CCGCTTTCATCCCGAGATGTGGGGACTCGAGGAACGACGACGAGCGCTCTCGCGCTCTC	5128
OY	511	TTCGACGCGCGCGCGACCGAGAAATCGCGGCATGCGGGAGACGTCGGCGAGGAAGCC	570
Db	5127	ATTTACAGGGGGCGCGCGCGAGCGCAATGTCTGGCTGACGAACCCGCGATGCAAGCC	5068
OY	571	ATGGCAGGCGCGCATTCGCAATTTGACACCGCGCGCTTACCCGCGCGCGCGCGCC	630
Db	5067	CCGACACGAGGCTTCAACCAAGTTCTGGCAACGCGCGGGTCCACGCGGCCCAACGGGCAAGGC	5008
OY	631	ACCACGAGAGATTCATCGAGGTGTGCGCGCGCTGAGCGCGCATGGCGGCAATCAGCC	690
Db	5007	TCTTTATTCACCGCGACCGCTGACCCCGCCCTCGCGACATCGGGGACGGCATGCGCC	4948
OY	691	ACCACATGTCGCGACGAAAGCGACATCGTGGCGCGCGCTGAGAGAACTTTCGGCATC	750
Db	4947	CAGAACCCGCTCTTGGGCCCTTCGCGCAGCGGCGGTGAGCGCGCTGGCCATGAGCGCGCTAC	4888
OY	751	G---GCCGAGGTGACGATGCGCGGTGTATCTCGACACACAAGATCATGGGCGCAAGCC	807
Db	4887	GATGCAACCCAGAAGACTTTCATCTCTCAGAGGCTTCGCGGGGCTTACGCGCATGAGCC	4828
OY	808	AATTTGGCGGCTCTGGCGAGACGCTGCGCTGATTCAGAGCGCGCATGCGCGCGCAAGAC	867
Db	4827	TACCGGGGCGCGCGCGGAGACCCCGCGCGGGAGGAGCGACCTGTGAGGCGCTCTCGCGCC	4768

[illegible]

```

RESULT 2
US-09-105-537-7
Sequence 7, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 1248
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-105-537-7

```

Query Match	4.73;	Score 81.8;	DB 4;	Length 1248;
Best Local Similarity	44.28;	Pred. No. 1.6e-07;		
Matches 439; Conservative	0;	Mismatches 542;	Indels 12;	Gaps 2

Qy	32	CCATGTCCTCCAAATCCGATTTCACAGCCCTTCGACTCTGCTGCGGGGAGACCTCATCG	91
Db	215	CCACCTGCAGACGACAGGCGCGGGGCTTCAGCTTCCTGCGACAGCGCGCGGCTTCACCGGCG	274
Qy	92	ACGGCAGACAAACACCCCGGGGGCGCGCGCGACCTGGGCGTCGCGGGGAGACGGCATTCGCG	153
Db	275	AAGTGATCATGCGCTGATGAGAGCTTCGCGCGCACCCGACGACATGCGCTGATGGGCG	334
Qy	152	CCATCGGCGCATCTGTCGAGCGCGCGCGGCGACACCCGGGTGCAGCTGTCGGGCGCTGGTG	211
Db	335	TCAACCCCGGCTTCGCGGAGCATGACCCGGAGACCGGACACTGCAGCCCGGAGACCAAGGTGG	394



```

QY 212 TCGGCGCCGCTTATCGACTGACACCCAGACAGCACTACCTGCTCAGCGCTGCG 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 395 CCGCGCGCGTACACACCCGACACCTGCGGCTGCTGCGGCTCCTCGGCGCGCCCT 454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 272 ACATGACGCCCAAGATCTCGAGGGGCTACACAGGCTGTCCAGCGGCAATTGCGCATCA 331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 455 GCGCGCGCGACCACTGCGGGAAGTCCGCCAGACAGACAGCGCTGCGGCTTCTGACG 514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 332 GCGTGGCGCGCTGCGGACGCAACCGCCCGCCCTGAGCTGCTGAGCAAGAGCG 391
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 515 CCGCGACAGCCCTCGGCTGCGGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCG 574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 392 GCTCTTACCGTTTCCAGGCGCTTCCGCGACTACGTGAGAGCGCTTGGCGGCGCACCGCGCG 451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 575 AGGTCTTACGTTTCA---CGCCACCAAGGCCGTCAAGCGCTTCCAGGGCGCGCGCTG 631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 452 CCGTCAACGCGCGCTGTATGTGTGCGCATTCACGCTGCGCGCGCGCTCATGCGGACT 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 632 TCACGAGAGAGCGCGACCTCGCGCGCGCGGATCCGCGCGCTCCACAACTTCCGCTTCCAGC 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 512 TCGACGCGCGCGCGCACCGACAGAAATCGCGCGCATCGCGGACTGCGCGAGAGCA 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 692 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 572 TGGCCAGCGCGCGCATTCGCGCATTTTCGACCGCGCGCTTTCACCGCGCGCGCGCGCG 631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 752 TGGCGCTTACCTCTCTCCAGCGCTTTCGAGGTATGATGACGCGAGCGCGCGCAACGACG 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 632 CCACGAGAGATCATGAGGTGTGCGCGCGCTGAGCGCGCATGCGCGCATTCACGCGCA 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 812 CCGCTACCGCGAGACCTCGCGCGCGCGCGCGCGCGCGCTCTGCGCGCGCGCGCGCG 871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 692 CCCACATGCGCGAGAGCGGACATCTGTGCGCGCGCTGAGAGAACTTCCGATCG 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 872 AGGCGCTCAACACCAACAGTACGTATGTCGAGATGACGACGCGCGCGCGCGCGCGCG 931
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 752 GCGCGAGGTGAGCGTGCCTGTGTATCTCGACACCAAGTCTATGAGCGACCGCAATY 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 932 ACCGCGACCTCTCATGAGGTCTGGAAGCGGAGGCGTGCACACCGCGCTTCTTCT 991
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 812 TCGCGCGCTGCGCGAGACGCTGCGCGTGTATGAG-----GCGCGCATGCGCGCGC 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 992 CGCGCGCGCTGCGCGAGGTGAGCGGTGTACCGCGCGCGCGCGCGCGCGCGCGCGCA 1051
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 863 AGGAGGTCTGCTGAGAGCGGTATCCCTACGTGCGCGCGCTCCACCATGCTCAAGAGAGACC 922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1052 CCGAAGCGCTCGCGCGCGCGCGCTGTCTCTCCGACGCGCGCGCGCGCGCGCGCGCG 1111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 923 GCGTGTCTGTGCGCGAGCGACCATCATCATCTGTGTCAGAGCGCTTCCCGAATGAGCG 982
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1112 ACATCGCGCGGCTCGCGCGAGCTGTGCTGTCTGTGCGCGACCGCGCGCGCGCGCGAG 1171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 983 GCGCGAGCTGTGATGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1172 CCGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCG 1204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 3
US-09-105-537-3/c
; Sequence 3, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0

```

```

; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-3

```

```

Query Match 4.7%; Score 81.8; DB 4; Length 13613;
Best Local Similarity 44.2%; Pred. No. 2.1e-07;
Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2.

```

```

QY 32 CCATGTCCCAATCGATTCGATCCAGCCCTTGCAGCTGTGCTGCGGCGGCGGACACCTTCATCG 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12669 CCACTGCAACGCGACGCGCGCGCGGCTCCAGTCTCTCGCGACGCGCGCGCGCTTCACGCGG 12610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 92 ACGGACAGAACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12609 AAGTATCATGCGGTGATGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 152 CCATGCGCGGATCTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12549 TCACCGCGGCTTGTGCGCGCATGACCCCGGACACCGGCAACTGTGACCGGACGAGTGG 12490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 212 TCGCGCGCGCGCTTATGACTGCGACACCGACGACGACGACGACGACGACGACGACG 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12489 CCGCGCGCGGTACACCCCGCGCACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 272 ACATGACGCGCGAGATCTGCGAGGCGCTCACACGAGTGTGACGCGGCAATGCGGATCA 331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12429 GCGCGCGCGAGACCTGCGGAGGTGCGCGAGAGACAGCGCGCGCGCGCGCGCGCTT 12370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 332 GCGTGGCGCGCTGCGCGACGCAACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 391
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12369 CCGCGCGACGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 392 GCTCTTACCGTTTTCGAGCGCTTGTGCGCGCATCTGACGCGGTTGCGGCGCGCGCG 451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12309 AGGTCTTACGTTTCA---CGCCACCAAGCGCGCTTCAACGCGCTTCAAGGCGCGCGCG 12253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 452 CCGTCAACGCGCGCGCTGTATGTTGGCGCATTCACACGCTGCGCGCGCGCGCGCGCG 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12252 TCACGAGAGAGCGCGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 512 TGCAGCGCGCGCGCGCGAGAGAAATCGCGCGCATGCGCGGACCTGCGCGAGAGAGCA 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12192 TCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 572 TGGCCAGCGCGCGCATGCGCATTTTCAGCGCGCGCGCTTTCACCGCGCGCGCGCGCG 631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12132 TGGCGCTTACCTCTCCGACGCGGTTTCCGAGTATGACACCGGACCGCGCGCGCG 12073
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 632 CCACGAGAGATCATGAGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12072 CCGCTTACCGCGAGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12013
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 692 CCCACATGCGCGAGAGAGCGAGACATCTGTGCGCGCGCGCTGAGAGAACTTCCGATCG 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12012 ACGGCTTCAACACCAACAGTAGATGATCTGCGAGATGACAGAGGCGCGCGCGCGCG 11953
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 752 GCGCGGAGCTGAGAGTCCCGGTGTATCTCGACCAACAAAGTATGAGCGCGCGCGCG 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11952 ACCCGACTTCTGTATGAGAGTCTGTGAAGCGCGAAGCGGTGACACCGCGCGCTTCT 11893
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 812 TCGCGCGGTGCGCGAGAGCGGTGCGGTGATGAG-----GCGCGCATGCGCGCGCG 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11892 CCGCGGCGTGCACAGACTGAGAGCGGTGACCGCGGCGCGCGCGCGCGCGCGCGCGCG 11833
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 863 AGGAGTCTGCTGAGAGCGGTATCCCTACGTGCGCGCGCTCCACCATGCTCAACGAGAGCG 922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11832 CCGAAGCGCTGCGCGCGCGCGGTGTCTGCTGCGCGACCGCGCGCGCGCGCGCGCG 11773
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 923 GCGTGTCTGTGCGCGGAGCGACCATCATCATCTGTGTCAAGCGCGCTTCCCGAATGAGCG 992
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11772 ACATTCGCGCGGTGCGCGAGCTGTGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 11713
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Oy 983 GGGCGCACTGGATGAATCTCGGGCCGAGCGC 1015  
 ||||| ||| | | | | | | | | |  
 Db 11712 CGCGCCACCGCAGACAGCGGCCCGCCGCTCG 11680

#### RESULT 4

```

US-09-773-816-1/c
: Sequence 1, Application US/09773816
: Patent No. 6340774.
:
: GENERAL INFORMATION:
: APPLICANT: Stanford University
: APPLICANT: Khosla, Chaitan
: TITLE OF INVENTION: NON-STEROIDAL ESTROGEN
: TITLE OF INVENTION: ANTAGONISTS
: FILE REFERENCE: 28600-20210.00
: CURRENT APPLICATION NUMBER: US/09/773,816
: CURRENT FILING DATE: 2001-06-08
: PRIOR APPLICATION NUMBER: US 60/243,458
: PRIOR FILING DATE: 2000-10-25
: PRIOR APPLICATION NUMBER: US 60/179,305
: PRIOR FILING DATE: 2000-01-31
: NUMBER OF SEQ ID NOS: 1
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 1
:
: LENGTH: 23673
:
: TYPE: DNA
:
: ORGANISM: Human
:
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(23623)
: OTHER INFORMATION: n = A,T,C or G
US-09-773-816-1

```

Query Match	Similarity	4.4%	Score 77.4	DB 4	Length 23673
Best Local	Similarity 42.4%		Pred No. 1.5e-06		
Matches 667	Conservative	0	Mismatches 891	Indels 15	Gaps

  

Oy	67	CTGCTCGGGGGCGGCACCTTCATCGACGGCAGCAGACACCCCGGGGGGGCGCGCGACCTG	126
Db	8742	CAGCTCCTGAGACCCCGGGCTCCCGCGAGCCGATCCGCAGACAGCTTCGACCTTCGACGTG	8683
Oy	127	GCGCTGCGGGCGCACCGGCATCGCCGCCATCGCGGATCTGTGGAGCCCGCCGGGCACAC	186
Db	8682	CTGCGCTTCAACCCACACCAAGCCCTTACCCCGCGCGCGCTGCGCCCGCGCTGGCGCCAC	8623
Oy	187	CGGCTGAGAGCTGCGGGCCGTGGTGGCGCGCCGGCTTCATCTGACATCGCACCCACAGAC	246
Db	8622	GGGCTCGACAGACCGCTGCTCATCTGTACCGCGCGATCCAGCCGACCGCTGTCTGTCTG	8563
Oy	247	GAAACATCACTGTGTCAAGGCTCGCGACATGACGCCCAAGATCTGCGAGGCGC--TCACC	303
Db	8562	GCCGCGATGACATTCGGGTTCCTCGGGGGAAGCTGGGGGGCGCGGCGGACCAATGATTC	8503
Oy	304	ACGGTGTCAACGGGCAATTCGGGCATCAGCTTGACCGCCCTGGGCGCACGGCAACCGGCC	363
Db	8502	ACCTTCGCCCCCGAGACCGCGCTCGGCCGAACACCTGCGCTGTGTATCTCACCGCGCTCC	8443
Oy	364	GCCCCCTTGGACCTGTGTGAGAGAAAGGGGGCTTTACCGTTTGAGAGCGCTTCGCGACATAC	423
Db	8442	GGCGCGCGCGGCATGACAGAAAGGGCGCTGCGCTCATGCAATGGCCAAAGACGACGCC	8383
Oy	424	CTGAGACCGTTTGGGGGCGACGCGCGGCGGCTCAACACCGCTGTATGTGGGCGCATTCA	483
Db	8382	GCCCTCGGGCAACTGTGAGAGAGCGCGGGGTGTGTGACATCTTCGTGGTACCGAGCCCAAC	8323
Oy	484	ACGCTGCGCGCGCGGTATCCGGACTTGTGACGCGCGCGCCACCGACAGAGAAATCGCG	543
Db	8322	TACGGCGCGCTCGGGCGTCTCTGCGACCTGTGTGCGACGTATGTCTGCGGAAACCGGC	8263
Oy	544	GCCATGCGGGACCTGTGCGAGAGAACCATGSCCAGCGCGCGCATTCGGATTTTCACCGGC	603
Db	8262	GCGGCTTCGCGCTTCGCGCGCGCGCGGCTCATTCGCGACAGACCATTCGCGAGAACTGCG	8203

Oy	604	GCCCTTACCGGCGCGCGCGCGGCACACACCAGAAAGATCATCGAGAGTGTCGGCGG	663
Db	8202	CGGAGTTCCAGACGCCGCCAATAACTCTCTCGAACACGGCCTGTATCACTGTATCGCG	8143
Oy	664	CTGAGCCGCGATGGCGGACATTACGCCACCCACATGCGCGACGAAGGCCAGCACATCTG	723
Db	8142	CGCGCGCGCTGGCGCGACGAACHTGGCCCCCTGCTGGCGCGCGCGCGAGCGGCCAAG	8083
Oy	724	CGCCGCTGGAAGAA---ACCTTCGGCATTCGGCCCGCGAGCTGAGAGTGGCGGTGATC	780
Db	8082	GGCGGGGACGGGGAACCGGGCCGACCGCGCGCGCTGAGACGGGGGGCATGCGCCTTACGTC	8023
Oy	781	TCCGACACAAGGTCATGGGCGCCAGCCCAATTTTCGGCCCTCGCGCGAGACGCTGCCGTCG	840
Db	8022	GTCGACCCCCECGACCTCGCGCGGGGACAGAACCCAGACCGGACAGGTGCCGGCGCGGGCC	7963
Oy	841	ATCGAGGCGCCCATGCGCGCGCCAGAGACTGTGCTGTGAGAGCGTATCCTTACGTGGCGCG	900
Db	7962	ATCGACCGGCGCCACACAGCTCTGACTACATCGCGCTGTGTGAGAGACTTGTGTGAACCTG	7903
Oy	901	TCCACACATGCTCAAAGAGAGACCGCGTGTGCTGGCGCGGACGACCATCATACCTGGTGC	960
Db	7902	CGCGGCGACCGCATCTGSGCGACTGCCCCCGCATGTGTGCGCGGCTGTGGCGCTTGGCG	7843
Oy	961	AAGCCCTTCCCGAATCGAGCGGGCGCGCGCATGTGATGAAGTGCGGCGGACCGCGCAAA	1020
Db	7842	GCGCGACCGCTGTGCGCGTATCGGCGACACAGAAGGGCCACAGACGCGCAGAGACTCGCGCG	7783
Oy	1021	TCCAAATACGACGTGTGTGCGCGAGACTGTACACCGCGCGCGCCATCTACTTCAATATGAC	1080
Db	7782	CCCAACTTCGGCGATG-----CGCATCCCTCGGGGTACCGGAAGCGCGCGGATCATG	7729
Oy	1081	GAAACCGACGTGACAGCATCTGTGGCGCGCCAGCATATGAGCTCGACGGCTGTG	1140
Db	7728	CGGCTTCGCGGACAAACTGTGTCTCGCGGTGTACCTTGTGTGACACCGCGCGCTTAC	7669
Oy	1141	CCGACACGAGACGCGCGGCATCGCGCTTGTGGGCACTTCCCGGSGTGTGGGGGAC	1200
Db	7668	CCCGCGCGGTGCGCGAGAGACAGAGGCGCAAGCGCTGTGCCATGCGCGAAGAACTGTGGCTG	7609
Oy	1201	TATGGGCGGACGTGGGCGCTTTCGCCCTGTGAGACGGCGGATGTAGAAATGACCGGCTG	1260
Db	7608	ATGGCGGCGCTGTGCGGTGTGACCTGTACCGTGTGTACCGCGGAGGGCGCGGCGGGG	7549
Oy	1261	AACGCGCGCGCTTGTGGCGCTGGCGGCGCGGCGAGCTGACGAGCGGG--TACTTCCGC	1317
Db	7548	CGCGTGTGCGCTGTGCGGTGCGCAACCGTGTCTGATGTTCAGAAACGGCATCTTACCTGGT	7489
Oy	1318	GACGTGTGTGTGTGACCGCGCGCGACAGGTGTGGCGATTAACCGCACCTTTCAGAACCTTAC	1377
Db	7488	ATTCAGCCCCGAGGGGTGGCGGCGCATCTGTGGGACGAGACCCCGCGCGCGCGGCTGCG	7429
Oy	1378	GAGCGCGCGCGCGGCATCATTTCCGTGTACGTCAACGCGCGCGCGGTGTGGCAAGAGAC	1437
Db	7428	GCCCGGCGCTTAGAGGTTCACGCGCGCGGGAACCTCTCGGCTGTGGCATGTGTACAGGCGTGC	7368
Oy	1438	GCGTTACCGCGCACATGCGCGCGCGGTGTGTGCACAGCACGGCGCGCTGTATCCCGGGCG	1497
Db	7368	CTTCCCGGAACCCGAAGGGGACACGGAAACCCGCTGTGGCGCGCGACGCGCTTCTCGC	7305
Oy	1498	CAGCCCTTACATTCGGGGGTGAACGGGGCGGCTGTGCGCGCCCTCTCCAAACCTTGTGAGCA	1557
Db	7308	GACGCGATCGCGCGCGGCTGTGGGAGATGTGTCCCTGTGACGGCCACAGATTTGGTGGCG	7244
Oy	1558	AACCGCTACATGGGCCCTCTCTCGCTGTCAATACGGCGCCACCCGATATCTGTGGCAAG	1617
Db	7248	GACCGCTGGGCGCGCTTCCGCGCTATGTGGCGGGAAGCACACAGCGCGCGAGATCAC	7185
Oy	1618	GAACTGATGGCGG	1630
Db	7188	GCCATGAAGGGG	7176



LOCATION: 36249..41774  
US-08-804-227C-1

Query Match 4.3%; Score 75.6; DB 2; Length 43280;  
Best Local Similarity 42.5%; Pred. No. 3.5e-06;  
Matches 722; Conservative 0; Mismatches 954; Indels 24; Gaps 5;

```

35 TGTCCAAATCCGATTCACACCCCTTCGACCTGCTGCTGCGGGGACACCTCCTACACG 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28810 TGGCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 GCAGCAACACCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28870 GGGTACACCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 TCGCGGATCTGTGCGACCCCGCCCGCCGACACCCCGGGGTGACGTGTGCGGCTGTGTGTCG 214
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28930 AAGAACCACTGGCACCGGGCCAGTACGATGCGCTCGGGGGGGGGGGGGGGGGGGGGGGGG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
215 CCGCGGGGTTCTGACATCTGACACACCCACAGACACCACTACCTGCTGAGGGGTGGCGACA 274
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28990 GCGAGCCCTGTATGCGCCCTCGGATGATACCGGGCA-----GGAAACATGGGGCGCG 29043
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
275 TGAACGCCAAGATCTGCGAGGGGCTACACAGGTGTCACGGGCAATTCGGGCATCAAGCC 334
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29044 AAGGAGCCGGGCTGCTGCTGAGACCGCCCGGATGTCACGGGCTTCGCGCGGAGAGAC 29103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
335 TGGCGCGCTGCGACACCCCAACCCCGCCCGCTGACCTGCTGACAGAGGGGCT 394
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29104 GCGTGTGCGCATGTGGAACGGCGGCTTCGGGCGCCCTGCTGCTGCGACACCGCATGG 29163
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
395 CTTCACGTTTCAGCGGCTTCGCGCACTACCTGAGCGGCTTGGGGGCAAGCGGGGGGGCG 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29164 TGGCCCCGATCCCGACGGCTGTGCTGACCGGAGGCGGCTTCGCGCGGCTGCTGCTGCTG 29223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
455 TCAACGCCCGCTGTATGTGAGGCCATTCAACGCTGCGCGCGCGGTCATGCGGACTTGC 514
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29224 TCACTCTCTACTGCGCGCTGACCGCGGCTGGCCCGGGCCGACCGGACAGACGCTCTG 29283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
515 AGCGCGCGCGACCGACAGGAAATGCGGGCATGCGGGGACCTGGCGGAGGAAAGCATGG 574
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29284 TGCACGCGCGCGCGGGGCTGTGCGATGGCGACCTTCACACTGCGCGGCACTGCGGC 29343
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
575 CCAAGCGCGCATGCGCATTTTCAGCCGGGCGCTTTCACCGCGCGCGCGCGCGCACCA 634
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29344 TGGAGGTGTACGCGACCGGACCGGCAAAATGGGACCGCTCGAGAGACAGCGATCC 29403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
635 CCGAAGATCATGAGGTGTGCGCGCGCTGAGCGGCTATGCGGACATCTACGCCACC 694
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29404 CCGACGACCGCATGCGGACTCCCGACCTG---GACTTCGCGGAGCGCTTCCTGTGTC 29460
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
695 ACATGCGCGGAGGAGGACATGCTGGCGCGGCTGGAGGAAACCTTCGATCGGCGC 754
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29461 GGAAGGCGCGCGGGGCTGTGACATGCTGTGAATCCTGCGCGCGGAGTTCTGACAG 29520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
755 GCGAGCTGAGCTGCGCGCTGTGTATTCGACACAGATGATGAGCCAGCCCAATTTG 814
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29521 CCTACATGCGGCTGTGCTGCGCGGGGCACTTCGGAATTCGCGAGCGGAGCTCC 29580
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
815 GCGCGCTGCGGAGACCTGCGCTGTATGAGCGCGCATGCGGCGCGGAGAGTCTCC 874
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29581 GCGAGCCCGCGGATGCGCGCGCCCATCGGGCACGACTACCGGGCTTCGACCTGG 29640
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
875 TGAAGCGGTATCCTAGTGGCGGCTTCACACATGCTCAAGAGAGACCGGCTGCTGG 934
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29641 TGAAGCGGCTGCGGACACCTCGGGGAGATGCTCGGGGAACTGTGGAATCTTGGCG 29700
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
935 CCGAGCGCATCATCATCTGTGCAAGCCCTTCCGAACTGAGCGGCGGAGCTGG 994
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29701 CCGAGCGCTGCGCGCGCTGCGCTACCGGCTACGCGATACGCGAGCGCGGAGC--- 29756
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
995 ATGAATGCGCGCGGAGGCGGCAATCAAGTCAAGTGTGCGGAGCTGAGCGCG 1054
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

29757 -----GCTTGGGACCCCTACAGGCGCGGACACCGGCAAGCTGCTGACGTCG 29811
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1055 CCGGCGCCATCTACTATGATGAGAGAACCCGAGCTGACAGCATCTGCGGCTGGCC 1114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29812 CTGCGCGATGTACACACCCGACGAGGTCTCTACCGGGGGGACGGGCGAGCTGGCC 29871
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1115 CGACATGATGCGCTCGAGGCTGCGCACAGAGAGCGCCGATCGCGGCTGTGG 1174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29872 AGACATCGCCCGCATCTGTGTACACCGGAGGGGTACGGGACCTGTGCTGGCGCC 29931
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1175 GCACCTTCCCGGGGTGTGGGCACTATGCGCGGACCTGGGCTGTTCCTCGTGAGA 1234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29932 GCACCGCGCGGCGCGGAGGGGTGTGCGGAATGATGCTGAGCGGAGTGGCG 29991
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1235 CCGCGGTATGAGATGATGACCGGCTGACCGCGCGGCTTTCGGGCTGGCG---GGCGCG 1291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29992 CCGAGGTGCGGCTGCGGCTGCGAGCGCGGACCGGAGCGGCTGACAGCTCTGCG 30051
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1292 GGCAGCTGACGCGCGGATCTTCCGACCTGCTGTGTGTGACACCGCGGACGCTGGCCG 1351
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
30052 CCGGAATCCCGCTGAGACACCGCTGCGCGCGCTGCTCCACCGCGGCGGACCTTCGACG 30111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1352 ATACCGCACCTTTCGAAACCTTACCGAGCGCGCGCGCATCTTCGCTGATGCTCA 1411
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
30112 ACAGCACCATCTCCCTACCTGACCGGAGAACATGCAACAGCTGCTGCGGCCAAGGCG 30171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1412 ACAGCGCGCGGCTGTGCAAGAGACCGCTTACCGGACGATGCGCGCGGCTGCTG 1471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
30172 ACAGCGCTGCTACACTGACAGACTGACCGGACCGGACCGGACCTCTGCGGCTGCTC- 30230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1472 CAGCAGCGCGCGCTGAGCGCGCGGCGGACCTTACATTCGCGGCTGAAACGGGCGGCT 1531
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
30231 --TACTGCTCTCTGCGGCTGCTGCGAGCGCGGACCGGCGGCTTACGCGCGGCA 30288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1532 GCGCGCGCTTCCCAACCTTGGAGGCAACCGCTTACATGAGCGGCTTCCCTCGCTGCAATA 1591
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
30289 ACAGCTTCTGTGAGGCTTGTGCGGATACCGGACGAGGCGCTGCGGCGCTTCCG 30348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1592 CCGCGCCACCGATATCTGTGGGCAAGAGATGATGGCGCGGCTGCGGCGGAGCGCA 1651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
30349 TGGCTGCGGAGCTGTGGGCGAGCAACAGCGCATGCGGCGGCGACCTGACCACTGCGGCA 30408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1652 AGCGCGGAAATATACCTTGAAGACTGTGCGAGCGGACGAGGATGCGGCTGCGAGCC 1711
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
30409 TGAACAGCGCGCTTACCGGAGCGGCAATGCGCTGACGAGCGGCGGCGCTGCGCC 30468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1712 TGTCCAGCGCGAGCTGGGC 1731
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
30469 TGTTCGAGCGCGGACAGGAC 30488
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 7  
US-08-461-775-8  
Sequence 8, Application US/08461775  
Patent No. 5858773  
GENERAL INFORMATION:  
APPLICANT: MAZODIER, Philippe  
APPLICANT: GUGLIELMI, Gerard  
TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE  
NUMBER OF INVENTION: INITIATION OF TRANSCRIPTION  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25









```

Db 4525 CGGCCCCCAAGCCCCCGGCGGAGGAGACCCGCGGCCCTTCCCGCCGCGCCG 4466
Qy 844 GAGGCCCCCAATGGCGGCGGAGAGCTCTGCTGGAGCCGATCCCTACGCGGCGCTCC 903
Db 4465 CACGCCCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4406
Qy 904 ACATGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 963
Db 4405 CTGCTGCTCATG-----CGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4352
Qy 964 CCGTTCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1023
Db 4351 GCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4292
Qy 1024 AAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1083
Db 4291 CGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4232
Qy 1084 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1143
Db 4231 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4172
Qy 1144 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1203
Db 4171 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4112
Qy 1204 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1263
Db 4111 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4052
Qy 1264 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1323
Db 4051 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3992
Qy 1324 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1377
Db 3991 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3932
Qy 1378 GAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1437
Db 3931 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3872
Qy 1438 GCGTTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1497
Db 3871 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3812
Qy 1498 CAGCCCTTACATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1541
Db 3811 TACTGCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3768

```

```

RESULT 12
US-09-105-537-1
Sequence 1, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
APPLICANT: Sherman, D. H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438051
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 15872
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-105-537-1

```

```

Query Match 4.2%; Score 73.8; DB 4; Length 15872;
Best Local Similarity 43.2%; Pred. No. 7e-06;
Matches 586; Conservative 0; Mismatches 747; Indels 24; Gaps 4;

Qy 175 GCGGCGACACCGGCGTCAACGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 234
Db 4799 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4858
Qy 235 CACACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 294
Db 4859 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4918
Qy 295 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 354
Db 4919 GACCTCACCGCGTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4978
Qy 355 AACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 414
Db 4979 CTGGAACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5038
Qy 415 GCGGAGTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 474
Db 5039 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 5098
Qy 475 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 534
Db 5099 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 5158
Qy 535 GAATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 594
Db 5159 GCGGAGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5218
Qy 595 TCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 654
Db 5219 GAGCCCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 5278
Qy 655 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 714
Db 5279 TGGCGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5338
Qy 715 CACATCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 774
Db 5339 CACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5398
Qy 775 GTGATCTCGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 834
Db 5399 CCGGCGGAGTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5451
Qy 835 CCGGCGGAGTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 894
Db 5452 --GCTCACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5509
Qy 895 GCGGCGTCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 954
Db 5510 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5569
Qy 955 TGTGCGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1014
Db 5570 GTCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5629
Qy 1015 GGCAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1074
Db 5630 GAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5689
Qy 1075 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1134
Db 5690 GGTCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5749
Qy 1135 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1194
Db 5750 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5809
Qy 1195 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1248

```

Query Match	4.2%;	Score 73.4;	DB 2;	Length 2167;
Best Local Similarity	46.7%;	Pred. No. 6.8e-06;		

100

REGISTRATION NUMBER: 36,113

REFERENCE/DOCKET NUMBER: 010830-035  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-2021  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2167 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-09-031-606-9

Query Match 4.2%; Score 73.4; DB 3; Length 2167;  
 Best Local Similarity 46.7%; Pred. No. 6.8e-06;  
 Matches 233; Conservativity 0; Mismatches 266; Indels 0; Gaps 0;

OY 297 CGTACCAAGGTGTGACGGGCAATTGGCGCATAGCCCTGGCGCCCTGGCGCCAA 356  
 DB 938 CGGCCCAAGGGCGGCAACGTCATCGACAGAAAGTTGCGCCCGCCGACCAACAA 997  
 OY 357 CCGGCGCGCCCGTGGACCTGCTGAGACGAGGGGCTTACCGTTTCGAGCGCTTGC 416  
 DB 998 CGAGGCGCTCACATCCCGCTGAGGTGAGTGGAGACCCGTCAGAGAACTCGGGC 1057  
 OY 417 CGACTACCTGACGGCTTGGGGCCACGCCGCGCTCAACGCGCGCTGTATGTGG 476  
 DB 1058 CCAGCTGCTAAGAGAGGTGGGACCAAGACCAACATCGCGGTGACGACCAACAC 1117  
 OY 477 CCATTCAACGCTGGCGCGCGCGGTGATCCGAGCTTGCAGAGCGCGCGCCGACGAG 536  
 DB 1118 CGGACCGCTGCTGGCGCGGCTGGTCCGGAGGGCTGCGCAAGTGGCGCGCGC 1177  
 OY 537 AATCGCGCATGCGGACCTGGCGGAGGAGCAATGGCCAGCGCGCTGATGCTTC 596  
 DB 1178 CTCGCCGCGCGCTGAGAGAGGATGAGCGCGCGCTGCGCGCTGCGCGGAGCT 1237  
 OY 597 GACCGCGCTTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 656  
 DB 1238 GCTGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 1297  
 OY 657 CGGCGCGCTGAGCGCGCATGGCGGCGATGACGCCAACCACATGGCGCGAGGAGG 716  
 DB 1298 CGCGGAGGACGAGAGGTGGGAGCTCATGCCGAGGAGGATGACAAAGTGGGAG 1357  
 OY 717 CATCTGCGCGCGCTGAGGAACCTTCGATCGCGCGCGGAGTGAAGCTGCGGT 776  
 DB 1358 CGGTGTATCAACGTGAGAGAGTCCACACCTTCGTTGACGCTGACCGAGGG 1417  
 OY 777 GATCTGCACCAAGATC 795  
 DB 1418 CATGCCCTTCGACAAAGGC 1436

RESULT 15  
 US-08-461-775-11  
 Sequence 11, Application US/08461775  
 Patent No. 5858773  
 GENERAL INFORMATION:  
 APPLICANT: MAZODIER, Philippe  
 APPLICANT: GUGLIEMI, Gerard  
 TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE  
 NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Burns, Doane, Swecker & Mathis  
 STREET: George Mason Bldg., Washington & Prince Sts.  
 CITY: Alexandria  
 STATE: Virginia  
 COUNTRY: United States  
 ZIP: 22313-1404  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/461,775  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/050,313  
 FILING DATE: 10-MAY-1993  
 APPLICATION NUMBER: FR 9011186  
 FILING DATE: 10-SEP-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Crane-Feury, Sharon E  
 REGISTRATION NUMBER: 36,113  
 REFERENCE/DOCKET NUMBER: 010830-035  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-2021  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2668 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-461-775-11

Query Match 4.2%; Score 73.4; DB 2; Length 2668;  
 Best Local Similarity 46.7%; Pred. No. 7e-06;  
 Matches 233; Conservativity 0; Mismatches 266; Indels 0; Gaps 0;

OY 297 CGTACCAAGGTGTGACGGGCAATTGGCGCATAGCCCTGGCGCCCGCTGGCGCCAA 356  
 DB 938 CGGCCCAAGGGCGGCAACGTCATCGACAGAAAGTTGCGCGCCCGCCGACCAACAA 997  
 OY 357 CCGGCGCGCCCGTGGACCTGCTGAGACGAGGGGCTTACCGTTTCGAGCGCTTGC 416  
 DB 998 CGAGGCGCTCACATCCCGCTGAGGTGAGTGGAGACCCGTCAGAGAACTCGGGC 1057  
 OY 417 CGACTACCTGACGGCTTGGGGCCACGCCGCGCTCAACGCGCGCTGTATGTGG 476  
 DB 1058 CCAGCTGCTAAGAGAGGTGGGACCAAGACCAACATCGCGGTGACGACCAACAC 1117  
 OY 477 CCATTCAACGCTGGCGCGCGGTGATCCGAGCTTGCAGAGCGCGCGCCGACGAG 536  
 DB 1118 CGGACCGCTGCTGGCGCGGCTGGTCCGGAGGGCTGCGCAAGTGGCGCGCGC 1177  
 OY 537 AATCGCGCATGCGGACCTGGCGGAGGAGCAATGGCCAGCGCGCTGATGCTTC 596  
 DB 1178 CTCGCCGCGCGCTGAGAGAGGATGAGCGCGCGCTGCGCGCTGCGCGGAGCT 1237  
 OY 597 GACCGCGCTTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 656  
 DB 1238 GCTGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 1297  
 OY 657 CGGCGCGCTGAGCGCGCATGGCGGCGATGACGCCAACCACATGGCGCGAGGAGG 716  
 DB 1298 CGCGGAGGACGAGAGGTGGGAGCTCATGCCGAGGAGGATGACAAAGTGGGAG 1357  
 OY 717 CATCTGCGCGCGCTGAGGAACCTTCGATCGCGCGCGGAGTGAAGCTGCGGT 776  
 DB 1358 CGGTGTATCAACGTGAGAGAGTCCACACCTTCGTTGACGCTGACCGAGGG 1417  
 OY 777 GATCTGCACCAAGATC 795  
 DB 1418 CATGCCCTTCGACAAAGGC 1436

Search completed: January 14, 2003, 06:36:00  
 Job time : 334 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: January 14, 2003, 08:03:39 ; Search time 2587 Seconds  
(without alignments)

11005.669 Million cell updates/sec

Title: us-10-009-782-1

Sequence: 1 gaattccactgacgcgcga.....ccctgacgacgagaacgtt 1758

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

EST:\*  
1: em\_estbta:\*  
2: em\_estbta:\*  
3: em\_estbta:\*  
4: em\_estbta:\*  
5: em\_estbta:\*  
6: em\_estbta:\*  
7: em\_estbta:\*  
8: em\_estbta:\*  
9: gb\_estbta:\*  
10: gb\_estbta:\*  
11: gb\_estbta:\*  
12: gb\_estbta:\*  
13: gb\_estbta:\*  
14: gb\_estbta:\*  
15: gb\_estbta:\*  
16: gb\_estbta:\*  
17: gb\_estbta:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_hum:\*  
20: em\_gss\_hum:\*  
21: em\_gss\_hum:\*  
22: em\_gss\_hum:\*  
23: em\_gss\_hum:\*  
24: em\_gss\_hum:\*  
25: em\_gss\_hum:\*  
26: em\_gss\_hum:\*  
27: em\_gss\_hum:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	83	4.7	1516	12 BG809984 mgct002xd
2	82.8	4.7	1798	17 AG171124 Pan trogl
3	82	4.7	1798	17 AG171124 Pan trogl
4	78.4	4.5	1542	17 AG032943 Pan trogl
5	76	4.3	1189	17 AG030608 Pan trogl
6	74	4.2	1116	12 BG810038 mgct002xd

7	72.4	4.1	925	17 CNS0091P	AL053013 Drosophi1
8	72.4	4.1	1299	17 AG039481	AG039481 Pan trogl
9	71.8	4.1	932	17 CNS00720	AL066742 Drosophi1
10	71.8	4.1	1341	17 AG030611	AG030611 Pan trogl
11	71.6	4.1	925	17 CNS0091P	AL053013 Drosophi1
12	71.2	4.1	1152	17 AG076818	AG076818 Pan trogl
13	71.2	4.1	1538	17 AG030607	AG030607 Pan trogl
14	71	4.0	1129	17 AG127788	AG127788 Pan trogl
15	71	4.0	1281	12 BG852363	BG852363 1024034A0
16	71	4.0	1288	14 B0678719	B0678719 AGENCOURT
17	70.2	4.0	1288	14 B0678719	B0678719 AGENCOURT
18	70.2	4.0	1131	17 AG042920	AG042920 Pan trogl
19	69.6	4.0	1625	17 AG043477	AG043477 Pan trogl
20	69.2	3.9	1360	13 BM463178	BM463178 AGENCOURT
21	69	3.9	1160	14 BM926731	BM926731 AGENCOURT
22	68.6	3.9	935	17 CNS006XK	AL066051 Drosophi1
23	68.6	3.9	1364	14 BM810045	BM810045 AGENCOURT
24	68.4	3.9	932	17 CNS00720	AL066742 Drosophi1
25	68.4	3.9	982	14 B0687717	B0687717 AGENCOURT
26	68.4	3.9	1626	10 AM731212	AM731212 GA_Ea001
27	68.2	3.9	1387	10 AM731212	AM731212 GA_Ea001
28	67.8	3.9	936	12 BG852371	BG852371 1024034A0
29	67.8	3.9	1232	17 AG072425	AG072425 Pan trogl
30	67.8	3.9	1328	17 AG043615	AG043615 Pan trogl
31	67.6	3.8	1088	17 AG073424	AG073424 Pan trogl
32	67.6	3.8	1299	17 AG039481	AG039481 Pan trogl
33	67.4	3.8	1189	17 AG162326	AG162326 Pan trogl
34	67.4	3.8	1189	17 AG030608	AG030608 Pan trogl
35	67.4	3.8	1278	17 AG060116	AG060116 Pan trogl
36	67.4	3.8	1651	12 BG809816	BG809816 mgct001xk
37	67.2	3.8	644	12 BG607300	BG607300 WHE2493_G
38	67.2	3.8	1248	17 AG049208	AG049208 Pan trogl
39	66.6	3.8	1040	12 BG786331	BG786331 SEAUMC006
40	66.4	3.8	1366	17 AG065058	AG065058 Pan trogl
41	66.2	3.8	1086	17 AG077581	AG077581 Pan trogl
42	66.2	3.8	1101	17 AG039543	AG039543 Pan trogl
43	66	3.8	1164	17 AG128373	AG128373 Pan trogl
44	65.8	3.7	1339	13 BM458211	BM458211 AGENCOURT
45	65.8	3.7	1452	17 AG032979	AG032979 Pan trogl

## ALIGNMENTS

RESULT 1  
LOCUS BG809984 1516 bp mRNA linear EST 22-MAY-2001  
DEFINITION mgct002xd11f Magnaporthe grisea Appressorium Stage cdna Magnaporthe  
grisea CDNA clone mgct002xd11f 5', mRNA sequence:  
BG809984  
VERSION BG809984.1 GI:14180964  
KEYWORDS EST.  
SOURCE Magnaporthe grisea.  
ORGANISM Magnaporthe grisea.

REFERENCE  
AUTHORS Sordariomyces Incertae sedis; Magnaportheaceae; Magnaporthe.  
TITLE 1 (bases 1 to 1516)  
JOURNAL Chol, W. and Dean, R.A.  
COMMENT Construction and sequence analysis of an appressorium stage CDNA library in the rice blast fungus, Magnaporthe grisea  
unpublished (2001)  
CONTACT: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: T3 primer (AATTAACTCTCACTAAAGCG).  
Location/Qualifiers  
1..1516  
/organism="Magnaporthe grisea"  
/strain="70-15"

FEATURES  
source

[illegible]

OY	948	CATPACTGTGGTGAAGCCCTTTCCCCGAACTGACGGGGCGGACCTGGATTGAATGTCGGGC	1007
Db	989	CGCGCGCCTGGCGGCGCCCGGCCCNCGCGCGCGCGCGGAGACCNCNGSAAACCCCGCGGCC	1048
OY	1008	CGAGCGGGGCAAAATTCAGAATACGACTGTGTGCCGAGACTGACGCGCGCGCATCTA	1067
Db	1049	GCCCG	1108
OY	1068	CTTGATGATGAGCAGAACCCGACGTGCAGCGCATCTTGCGCTTGCGGCCGACCATGATCG	1127
Db	1109	GNGCCCG	1168
OY	1128	CTTCGAGGCGCTTGGCGCGCGAGCAGAGCGCGCGCATCTCGCGCGCTGTGGGACACTTCCGCG	1187
Db	1169	CG	1228
OY	1188	GGTGTGGGCGCACTATGCGCGCGCACCTGTGGCTTCTCCGTGAGACGCGGTATGAA	1247
Db	1229	ACNCGAGCGGGGAGNCG	1288
OY	1248	GATGACGCGCGCTGACCGCGCGCGCGCTTGCGGCTTGCGCGCGCGCGCGCGCGCGCG	1307
Db	1289	GCG	1348
OY	1308	GTACTTCCCGCACCTGTGTGTGTTCGACCCCGCGCACGCGTGGCGCGATACCGCGCATTTGCA	1367
Db	1349	CCGCACACG	1408
OY	1368	ACACCTTACCGAGCGCGCGCGCGCGCATTCATTCGCTGAC	1408
Db	1409	CG	1449
<hr/>			
RESULT 2			
LOCUS	AG171124	1798 bp	DNA linear GSS 09-JAN-2002
DEFINITION	Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey		
ACCESSION	AG171124		
VERSION	AG171124.1	GI:16700802	
KEYWORDS	GSS.		
SOURCE	Pan troglodytes male lymphocytes DNA, clone_1lb:RPCI-43 Chimpanzee Pan troglodytes Male BAC library clone:RP43-040F09.TJ. Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.		
ORGANISM	Pan troglodytes		
REFERENCE	1 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tochki,Y., Watanabe,H. and Sakaki,Y. BAC end sequences of Library RPCI-43 Unpublished 2 (bases 1 to 1798) Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tochki,Y., Watanabe,H. and Sakaki,Y. Direct Submission Submitted (02-AUG-2001), Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan 1-7-22 Suheiho-chou,Tsukumi-kp, Yokohama, Kanagawa 230-0045, Japan (E-mail:shueibgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)		
JOURNAL			
AUTHORS			
TITLE			
COMMENT	Clones are derived from the chimpanzee BAC library RPCI-43 This BAC end was generated during the Rad process and may have higher chances of clone tracking errors. PRIMERS Sequencing: TJ LIBRARY Vector : pBACE3.6 R.Site 1 : EcoRI R.Site 2 : EcoRI Location/Qualifiers 1..1798 /organism="Pan troglodytes" /db_xref="taxon:9598"		











[illegible]

Accession	LOCUS	DEFINITION
U00001	CNS0091P	925 bp DNA linear GSS 03-JUN-1999
U00002	CNS0091P	Drosophila melanogaster genome survey sequence TE13 end of BAC #
U00003	CNS0091P	BACR19D16 of RP11-98 library from Drosophila melanogaster (fruit
U00004	CNS0091P	fly), genomic survey sequence.
U00005	CNS0091P	U000012

ACCESSION VERSION KEYWORDS SOURCE ORGANISM	AL053013 AL053013.1 GI:4934461 GSS. <i>Drosophila melanogaster</i> . <i>Drosophila melanogaster</i> . Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; <i>Drosophila</i> . 1 (bases 1 to 925)
REFERENCE	

RECEIVED  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 925)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaton Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPci-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES	Location/Qualifiers
source	1..925
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone="BACR1D16"
	/clone_1lb="RPCI-98"
	/note="end : TET3"
BASE COUNT	120 a 61 c 61 g 172 t 511 others
ORIGIN	

```
Query Match          4.1%   Score 72.4 : DB 17; Length 925;  
Best Local Similarity    13.3%; Pred. NO. 0.017;  
Matches      51; Conservative 190; Mismatches 14; Indels 0; Gaps 0;
```

QY	1152	GCGCCGCATCCGCGACCTGTGGGCGACCTTCCCGGGGTGTGGGCGACATGACGCGCGA	1211
Db	602	SSBSSSSKCSTBSKSCCCCKSVCTSCSSSSSSSSSSSTSSSTSKSSSSGS	661
QY	1212	CTTGCGCCCTGTCCCGCTGCAGACGGCGGTATGAGAATGACCGCGCTGACCGCGCGC	1271
Db	662	SSSSSSSYTTATSKSTBSAGSGSWSAGGSGSGTSGTSSSSSSSTSTSSSVSGKSTBS	721
QY	1272	CTTGCGCGCTGCACGGCGCGGACGTGACAGCGCGGTACTTCCGACCATGTGGGTGT	1331
Db	722	SGBSSSSGSSSSSTSSBSCTSTSSSSSSSYSSSTSCCTCCCSYSSSTSSSTSM	781
QY	1332	CGACCCGGCCACGGTGGCCCATACCGCACCTTGGAAACCCCTTACCGACGGCGCGCCGG	1391
Db	782	GSTGSSSSSYGTSSSSDSTSTCCSCCYMCTCSTYBMDEYTSISGCGSSSSGKGVT	841
QY	1392	CATCATCATCCCTGTACGCTCAACGGCGCGCGGTGTGCACAGACAGCGTTCACCGGCA	1451
Db	842	KCGCGGCGSSSTNGBMBSACSSSSSSSSSSSVSSSKSABSSSVSSSSSGCVSNSS	901
QY	1452	GCATCCGCGCGCGCTGCTGCACG	1475
Db	902	SASKSSSGSVSGSGSGSVSG	925

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE
RESULT 8	AG039481	1299 bp DNA	linear	GSS 01-NOV-2001		
	AG039481	Pan troglodytes DNA, clone: PTB-016J15.F, genomic survey sequence.				
	AG039481	Pan troglodytes DNA, clone: PTB-Chimpanzee Male				
	AG039481.1	GI:16568206				
	GSS.	Pan troglodytes male lymphoblast DNA, clone: lib:PTB				

ORGANISM	Pan troglodytes Eukaryotes: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE	1
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Motoki, Y., Watanabe, H., and Sakaki, Y. BAC end sequences of library pMB
TITLE	unpublished
JOURNAL	2 (bases 1 to 129)
REFERENCE	1
AUTHORS	Watanabe, M., Toyoda, A., Fujiyama, A., Hattori, M., Yada, T., Motoki, Y., Sakaki, Y., and Taylor, T. D.

TOTOKI, Y., MATANABE, H. and SAKAKI, Y.  
 Direct Submission  
 Submitted (02-NOV-2001) Asao Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
 1-7-72 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail: chimpbes@gsc.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/>,  
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
 Clones are derived from the chimpanzee BAC library PTB This BAC end  
 was generated during the R&D process and may have higher chance of  
 clone tracking errors.  
 PRIMERS

```

LIBRARY
  Vector      : PKSI45
  R.Site 1    : Saci
  R.Site 2    : Saci
  Location/Qualifiers
    1. 1299
FEATURES
  source

```

Query Match	4.1%	Score 72.4	DB 17	Length 1299
BASE COUNT	24 a	571 c	592 g	31 t
ORIGIN	//sex="male" /cell_type="lymphoblast" /clone_id="PTB Chimpanzee Male BAC Library" others			





```

ACCESSION   AC030611
VERSION     AC030611.1
KEYWORDS    GI:16557484
SOURCE      Pan troglodytes male lymphoblast DNA, clone_11b.PTB Chimpanzee Male
            BAC Library clone:PTB-003A20.F.
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       BAC end sequences of library PTB
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 1341)
AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
            E-mail:chimpdes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
            Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
COMMENT     PRIMERS
            Sequencing: -21M13
LIBRARY     Vector : PKS145
            R.Site 1 : SacI
            R.Site 2 : SacI.
FEATURES    location/Qualifiers
            1..1341
            /organism="Pan troglodytes"
            /db_xref="taxon:9596"
            /clone="PTB-003A20.F"
            /sex="male"
            /cell_type="Lymphoblast"
            /clone_11b="PTB Chimpanzee Male BAC Library"
BASE COUNT  48 a 476 c 700 g 14 t 103 others
ORIGIN
Query Match
Best Local Similarity 44.1%; Score 71.8; DB 17; Length 1341;
Matches 245; Conservative 0; Mismatches 307; Indels 3; Gaps 1;
QY 1105 GCGTTGCGCCGACCGACGATGCTCGAGCGCTCCGCGCAGCAGAGCGCCGATCGG 1164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1165 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 1106
QY 1165 GCGCTGCGGCGACCTTCCCGCGGCTGCGGACATATGCGCGCGACCTGCGCTGTC 1224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1105 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 1046
QY 1225 CCGGTGAGAGCGGCGGTATGAGATGACCGGCTGACCGCGCGCTTTCGGCTGACC 1284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1045 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 986
QY 1285 GGGCGCGGACGCTGCAGCGCGGCTACTTCCGCACTGTGTGTTCGACCCGCGCAG 1344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 985 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 926
QY 1345 GTGGCGGATACCGGACCTTGAACACCTTACCGAGCGCGCGCGCGCGGATCCATTCG 1404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 925 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 866
QY 1405 TACGTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1461
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 865 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 806
QY 1462 GCGGTGCTGCAGCAGCGCGCGCGCTGAGCCCGCGCGCGCGCGCGCGCGCGCGTGA 1521
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 805 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 746

```

```

QY 1522 GGGCGCGGCTGCCCGCCCTCTCCCAACCTGAGACGCAAAACGCTACATGCCCCCTCC 1581
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 745 NNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 686
QY 1582 GCGCGCAATRAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 685 CCGNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 626
QY 1642 GCGGACGCGCAAGCGCC 1656
    ||| ||| |||
Db 625 GCGCGCGCGCGCGCGCC 611
RESULT 11
CNS0091P/c
LOCUS
DEFINITION  Drosophila melanogaster genome survey sequence TEP3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1
KEYWORDS    GI:4934461
SOURCE      Drosophila melanogaster.
ORGANISM    Drosophila melanogaster.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Prexygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 925)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage;
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutyo Osoegawa and
            Aaron Mammosser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES    location/Qualifiers
            1..925
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="BACR19D16"
            /clone_11b="RPCI-98"
            /note="end : TEP3"
BASE COUNT  120 a 61 c 61 g 172 t 511 others
ORIGIN
Query Match
Best Local Similarity 14.0%; Score 71.6; DB 17; Length 925;
Matches 55; Conservative 189; Mismatches 150; Indels 0; Gaps 0;
QY 1134 GCGCTCCCGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 925 CSBSCSCSCSSBSSTSSMSSTSSMSSTSSMSSTSSMSSTSSMSSTSSMSSTSSSS 866
QY 1194 GGGGCACTATGCGGCGACCTGCGGCTTCCCGGTGAGACGCGCGGTATGAGATGAC 1253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 865 SSGTSSACVACNMSSSCGCGCGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 806
QY 1254 CCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTACT 1313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 805 GGASASHSBSSSSSSSASCSWMSASSSSSRSGGAGSGASSSSSSSS 746  
QY 1314 CGCGACCTGTGTGTTTCACCGCGGCGGATACCGCATCTTCAACACC 1373  
Db 745 ASAGSVYSSASSSSSSSVSCSVASSMSGSRSSSASASSSSSSSASCASC 686  
QY 1374 TACCGAGCGCGCGCGCATTCCTGTGTCACGCGCGCGCGTGGCAGA 1433  
Db 685 TMSGCSCTASMSASASSSSSSSSSSSSSSSSSSSSSSSSSSSSSSACB 626  
QY 1434 GCAGGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1493  
Db 625 SMSGGGSVSSASSSSSVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 566  
QY 1494 GCAGCAGCCCTTACATCCGCGCGTGAACGGGCG 1527  
Db 565 SGCWCRCSSAAASAAASVAAASGMMGCKSKSG 532

RESULT 12  
AG076818 1152 bp DNA linear GSS 03-NOV-2001  
LOCUS Pan troglodytes DNA, clone: PTB-071C05.R, genomic survey sequence.  
DEFINITION AG076818  
ACCESSION AG076818  
VERSION AG076818.1 GI:16628620  
KEYWORDS GSS.  
SOURCE Pan troglodytes male lymphoblast DNA, clone.lib:PTB Chimpanzee Male  
ORGANISM BAC library clone:PTB-071C05.R.  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE  
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE BAC end sequences of library PTB  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 1152)  
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:chimpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.  
PRIMERS  
Sequencing: M13Rev  
LIBRARY  
Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI.  
FEATURES  
source  
1.1152  
Location/Qualifiers  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-071C05.R"  
/sex="male"  
/cell\_type="Lymphoblast"  
/clone.lib="PTB Chimpanzee Male BAC library"  
BASE COUNT 34 a 488 c 514 g 23 t 93 others  
ORIGIN

Query Match 4.1%; Score 71.2; DB 17; Length 1152;  
Best Local Similarity 45.3%; Pred. No. 0.028;  
Matches 308; Conservative 0; Mismatches 368; Indels 4; Gaps 3;

QY 105 CCGCGGCG 164  
Db 389 CG 448

QY 165 GTCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 224  
Db 449 GCG 508  
QY 225 CATGCACTGCAACACCCAGACGACAATCACTGCTCAGGCTCGGACATGACCCAA 284  
Db 509 GGGGCGGCG 566  
QY 285 GATCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 344  
Db 567 GGGGCG 626  
QY 345 GCGCGACCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 404  
Db 627 GGGGCG 686  
QY 405 CGAGCGCTTCCG 463  
Db 687 CCG 746  
QY 464 CCTGTATGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 523  
Db 747 GCG 806  
QY 524 CCACCGACGAGGAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 583  
Db 807 CGGCG 866  
QY 584 CCATGCGCATTTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 643  
Db 867 GCG 925  
QY 644 TCATGAGGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 703  
Db 926 GCG 985  
QY 704 ACAGAGCG 763  
Db 986 GCG 1045  
QY 764 ACCTGCGGTGTGATCTCG 783  
Db 1046 CCG 1065

RESULT 13  
AG030607 1538 bp DNA linear GSS 01-NOV-2001  
LOCUS Pan troglodytes DNA, clone: PTB-003A18.F, genomic survey sequence.  
DEFINITION AG030607  
ACCESSION AG030607  
VERSION AG030607.1 GI:16557480  
KEYWORDS GSS.  
SOURCE Pan troglodytes male lymphoblast DNA, clone.lib:PTB Chimpanzee Male  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE  
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE BAC end sequences of library PTB  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 1538)  
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:chimpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the R&D process and may have higher chance of

## Clone tracking errors:

## PRIMERS

Sequencing: -21M13

## LIBRARY

Vector : PKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. 1538

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone="PTB-003A18.F"

/sex="male"

/cell\_type="lymphoblast"

/clone.lib="PTB Chimpanzee Male BAC Library"

BASE COUNT 67 a 669 c 552 g 64 t 186 others

## ORIGIN

Query Match 4.1%; Score 71.2; DB 17; Length 1538;

Best Local Similarity 39.7%; Pred. No. 0.03;

Matches 422; Conservative 0; Mismatches 634; Indels 6; Gaps 4;

QY 612 CCGGCCCCCGCCCGCCACACCGAAGATCATGAGTGTGCGCGCGTACGCGC 671  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 390 CCGGCGGGGCGCCCGCCCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 449  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 672 GCATGGCGGCATGTACGCGCACCATGCGGCGAGAGGCGGACACATCGTGGCGCGCT 731  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 450 GGGGCGGCG 509  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 732 GGAGAAACCTTC-CGCATCGCGCGCGAGCTGAGCTGCGGTGTGATCTCCACACA 790  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 510 GCG 569  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 791 AGGTGATGAGCGCGACCAATTTGGCGCGCTCGCGCGCGAGACGCTCGCTGATGAGCGCG 830  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 570 CGGCG 629  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 851 CCATGGCGCGCGCGAGAGCTGTGCGTGTGAGCGCTATCCCTAGTGGCGCGCGTCCACCATGC 910  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 630 CCGGCG 669  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 911 TCAGCAGAGACCGCGCTGCTGCGCGCGAGCAGCATCATCATCTGCTGCAACGCTTTC 970  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 690 GCG 749  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 971 CCGAAGTGAAGCG 1030  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 750 C--GGCG 806  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1031 ACCTGTGCGCGCGAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1090  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 807 CCGGCG 866  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1091 TGCAGCGCATCTGCGGCTTTCG 1150  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 867 CCGGCG 925  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1151 AGGCGCGCATCTGCGCGCTTTCG 1210  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 926 GCGGCG 985  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1211 ACCTGGCGCTGTTCGCGCTGAGACGCGCGGTATGAGAGATGACCGCGCTGACCGCGCGCG 1270  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 986 ACACG 1045  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1271 GCTTTCGCGCTGCG 1330  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1046 GCG 1105  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1331 TCGACCG 1390  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1106 CGACG 1165  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1391 GCATCCATTCGCTGATGCTCAACGCGCGCGCGGTGTGCGAGAGAGCGCTTACCGCGC 1450  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1166 GCG 1225  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1451 AGCATGCG 1510  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1226 CCGGCG 1285  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1511 CCGGCGTGAACG 1570  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1286 GC-GCAGCG 1344  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1571 CCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1630  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1345 CCGGCG 1404  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1631 CCGGCG 1672  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1405 CCGGCG 1446  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14  
 AG127788/c 1129 bp DNA Linear GSS 04-NOV-2001  
 LOCUS AG127788/c  
 DEFINITION Pan troglodytes DNA, clone: PTB-138M02.R, genomic survey sequence.  
 ACCESSION AG127788  
 VERSION AG127788.1 GI:16656953  
 KEYWORDS GSS.  
 SOURCE Pan troglodytes male lymphoblast DNA, clone.lib:PTB Chimpanzee Male  
 BAC Library clone:PTB-138M02.R.  
 ORGANISM Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE  
 AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
 Totoki,Y., Watanabe,H. and Sakaki,Y.  
 TITLE BAC end sequences of library PTB  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1129)  
 AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
 Totoki,Y., Watanabe,H. and Sakaki,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
 1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:chimpesgsc.riken.go.jp, URL:http://bgp.gsc.riken.go.jp/,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 Clones are derived from the chimpanzee BAC library PTB This BAC end  
 was generated during the Rad process and may have higher chance of  
 clone tracking errors.

COMMENT  
 PRIMERS  
 Sequencing: M13rev

LIBRARY  
 Vector : PKS145  
 R.Site 1 : SacI  
 R.Site 2 : SacI  
 Location/Qualifiers  
 1. 1129  
 /organism="Pan troglodytes"  
 /db\_xref="taxon:9598"  
 /clone="PTB-138M02.R"  
 /sex="male"  
 /cell\_type="lymphoblast"  
 /clone.lib="PTB Chimpanzee Male BAC Library"

BASE COUNT 54 a 369 c 480 g 44 t 182 others

ORIGIN  
 Query Match 4.0%; Score 71; DB 17; Length 1129;  
 Best Local Similarity 38.3%; Pred. No. 0.03;  
 Matches 352; Conservative 0; Mismatches 565; Indels 3; Gaps 2;  
 QY 822 GCGGAGAGCGCTGCG 881  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||









transformed microorganism from zinc-tolerant host for selective  
production of D-aminoacylase, useful in synthesis of high  
optical-purity D-amino-acids for antibiotic side-chains and peptide  
drugs -

Claim 2: Page 14-18; 22pp; Japanese.

The invention relates to a recombinant zinc-tolerant microorganism which  
expresses the D-aminoacylase from Alcaligenes xylosoxidans subspecies  
xylosoxidans. The presence of zinc ions in the culture medium enhances  
expression of the D-aminoacylase gene, and the invention also relates to  
the process of recombinantly producing the D-aminoacylase using the  
microorganism of the invention. The recombinant microorganism is used  
for the selective production of D-aminoacylase, which is useful in the  
synthesis of high optical-purity D-form amino acids for use in the  
production of antibiotics and peptide drugs. The present sequence  
represents DNA encoding D-aminoacylase from Alcaligenes xylosoxidans  
subspecies xylosoxidans.

Sequence 1758 BP; 281 A; 662 C; 568 G; 247 T; 0 other;

Query Match 100.0%; Score 1758; DB 22; Length 1758;

Best Local Similarity 100.0%; Pred. No. 5.8e-267;

Matches 1758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 GAATTCACCTTATGCGGAGAGAGATTTCATGTCCTCCATCCGATTCACGCTTC 60
DB 1 GAATTCACCTTATGCGGAGAGAGATTTCATGTCCTCCATCCGATTCACGCTTC 60
OY 61 GACCTGCTCTCGCGGGGCGACCTCATGACGCGACAGACACCCCGGGGCGCGCC 120
DB 61 GACCTGCTCTCGCGGGGCGACCTCATGACGCGACAGACACCCCGGGGCGCGCC 120
OY 121 GACCTGCGGCGGCGGCGACCGCATCGCGGCGATCGTGGAGCGCGCGCGG 180
DB 121 GACCTGCGGCGGCGGCGACCGCATCGCGGCGATCGTGGAGCGCGCGCGG 180
OY 181 CACACCGGGGTGCGACTGTGCGGCGCTGTGTGCGCGCCCGCTTATCGACTCGACACC 240
DB 181 CACACCGGGGTGCGACTGTGCGGCGCTGTGTGCGCGCCCGCTTATCGACTCGACACC 240
OY 241 CACGACGACACTACCTGCTCAGGCGTCCGACATGACGCCCAAGATCTCGAGGGCGTC 300
DB 241 CACGACGACACTACCTGCTCAGGCGTCCGACATGACGCCCAAGATCTCGAGGGCGTC 300
OY 301 ACCAGGCTGTCAGGCGCATTCGCGCATCAGCTGCGCGCGGTGCGACGCAACCGG 360
DB 301 ACCAGGCTGTCAGGCGCATTCGCGCATCAGCTGCGCGCGGTGCGACGCAACCGG 360
OY 361 CCCGCCCCCTGGACCTGTGACGAGAGCGGCTTTACCGTTTGCAGCGCTTCCGCGAC 420
DB 361 CCCGCCCCCTGGACCTGTGACGAGAGCGGCTTTACCGTTTGCAGCGCTTCCGCGAC 420
OY 421 TACCTGAGAGCGGTTGGGGGCGACGCGGCGGCGTCAAGCGCGCTGTATGGTGGGCGAT 480
DB 421 TACCTGAGAGCGGTTGGGGGCGACGCGGCGGCGTCAAGCGCGCTGTATGGTGGGCGAT 480
OY 481 TCAACGCTGCGCGCGCGGTCATGCGGACTTTCAGCGCGCGCGCACGACGAGAAATC 540
DB 481 TCAACGCTGCGCGCGCGGTCATGCGGACTTTCAGCGCGCGCGCACGACGAGAAATC 540
OY 541 GCGGCGATGCGGAGACCTGCGGAGAGAACCATGCGACGCGCGCGCATTCGACCC 600
DB 541 GCGGCGATGCGGAGACCTGCGGAGAGAACCATGCGACGCGCGCGCATTCGACCC 600
OY 601 GGGGCTTTCATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
DB 601 GGGGCTTTCATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
OY 661 CCGCTGAGCGCGCATGCGCGCATTCAGCCACCCACATGCGGCGAGGCGAGCATC 720
DB 661 CCGCTGAGCGCGCATGCGCGCATTCAGCCACCCACATGCGGCGAGGCGAGCATC 720
```

```

OY 721 GTGGCCGCGCTGAGAGAAACCTTCCGATCGCGCGGAGCTGAGACGTGCGGATGATC 780
DB 721 GTGGCCGCGCTGAGAGAAACCTTCCGATCGCGCGGAGCTGAGACGTGCGGATGATC 780
OY 781 TCGCACACAAAGTCTATGCGCGACCCCAATTTGGCGCGCTCGCGGACAGCGTCCGCTG 840
DB 781 TCGCACACAAAGTCTATGCGCGACCCCAATTTGGCGCGCTCGCGGACAGCGTCCGCTG 840
OY 841 ATGAGCGCGCATGGGCGCGGACGATCGCTGAGCGCGATTCCTACGAGCGCGG 900
DB 841 ATGAGCGCGCATGGGCGCGGACGATCGCTGAGCGCGATTCCTACGAGCGCGG 900
OY 901 TCCACATGCTCAAGACAGAGACCGGCTGCTGCGCGGACGACCATCATCACCCTGTGTC 960
DB 901 TCCACATGCTCAAGACAGAGACCGGCTGCTGCGCGGACGACCATCATCACCCTGTGTC 960
OY 961 AAGCCCTTCCCGGACACTGAGCGGGCGGACCTGATGATAGTTCGCGCGGCGGCGAAA 1020
DB 961 AAGCCCTTCCCGGACACTGAGCGGGCGGACCTGATGATAGTTCGCGCGGCGGCGAAA 1020
OY 1021 TCCAGTACGAGCTGTGTCGCCGAGCTGACGCGCGCGGCGCATCTACTCATGATGAC 1080
DB 1021 TCCAGTACGAGCTGTGTCGCCGAGCTGACGCGCGCGGCGCATCTACTCATGATGAC 1080
OY 1081 GAACCCGACGTCAGACGCGCATCTGCGGCGTTCGCGCGGACATGATGCGGCTCGACGCGCTG 1140
DB 1081 GAACCCGACGTCAGACGCGCATCTGCGGCGTTCGCGCGGACATGATGCGGCTCGACGCGCTG 1140
OY 1141 CCGGACAGACAGCGCGCGCATCTGCGGCGCTGTGCGGCGACCTTCCCGCGGCTCGGCGAC 1200
DB 1141 CCGGACAGACAGCGCGCGCATCTGCGGCGCTGTGCGGCGACCTTCCCGCGGCTCGGCGAC 1200
OY 1201 TATGCGCGGACCTGTGCGGCTTCCCGCTGAGAGAGCGGCTATGAGATGACCGGCTG 1260
DB 1201 TATGCGCGGACCTGTGCGGCTTCCCGCTGAGAGAGCGGCTATGAGATGACCGGCTG 1260
OY 1261 ACCGCGCGCGCTTGGCGCTGCGGCGGCGGCGGACGCTGACGCGCGGCTTTCGCGAC 1320
DB 1261 ACCGCGCGCGCTTGGCGCTGCGGCGGCGGCGGACGCTGACGCGCGGCTTTCGCGAC 1320
OY 1321 CTGGTGGTTCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380
DB 1321 CTGGTGGTTCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380
OY 1381 CGCGCGCGCGCATCTCGTGTACGTCAACGCGCGCGGCTGTGGCAAGACAGGCG 1440
DB 1381 CGCGCGCGCGCATCTCGTGTACGTCAACGCGCGCGGCTGTGGCAAGACAGGCG 1440
OY 1441 TTCACCGGCGACATGCGGCGCGGCTGCTCGGACGACGCGCGGCTGAGCGCGCGCGAC 1500
DB 1441 TTCACCGGCGACATGCGGCGCGGCTGCTCGGACGACGCGCGGCTGAGCGCGCGCGAC 1500
OY 1501 CCTTACATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
DB 1501 CCTTACATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
OY 1561 CGCTACATGCGCGCTTCCCTCGCTGCGCAATAGCGCGCGCACCGCATGTGTGGCAAGAA 1620
DB 1561 CGCTACATGCGCGCTTCCCTCGCTGCGCAATAGCGCGCGCACCGCATGTGTGGCAAGAA 1620
OY 1621 GTGATGCGCGCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680
DB 1621 GTGATGCGCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680
OY 1681 TCGAGGCGGCGGCGGATGCGGCTGTGACGCGCTGTCAAGGCGGAGCTGGGCGAGTCCG 1740
DB 1681 TCGAGGCGGCGGCGGATGCGGCTGTGACGCGCTGTCAAGGCGGAGCTGGGCGAGTCCG 1740
OY 1741 CTGAGCTACGAGAGCTT 1758
DB 1741 CTGAGCTACGAGAGCTT 1758
```

XX	AAAI5140		standard; DNA; 1473 BP.
XX	AAAI5140;		
XX	21-AUG-2000	(first entry)	
XX	DNA encoding an aminocyclase enzyme.		
XX	Aminocyclase; N-acetyl-D-tryptophan; R-N-acetyl-2-thienylalanine;		
KW	R-N-acetyl-4-chlorophenylalanine; D-amino acid; N-acyl amino acid;		
KW	pesticide; antibiotic; ss.		
OS	Alcaligenes sp.		
FH	Key	Location/Qualifiers	
FT	CDS	7..1461	
FT	/product=	"aminocyclase"	
PN	WO200023598-A1.		
XX	27-APR-2000.		
XX	20-OCT-1999;	99WO-GB03458.	
XX	20-OCT-1998;	98GB-0022947.	
PR	01-APR-1999;	99GB-0007739.	
PA	(CHIR-) CHIROTECH TECHNOLOGY LTD.		
XX	Taylor StC.	Brown RC;	
DR	WPI: 2000-336997/29.		
DR	P-PSDB; AAY84943.		
PT	New isolated D-amino acylase enzyme, useful for the preparation of		
PT	D-amino acids for use as intermediates in the preparation of		
PT	pesticides, antibiotics and other pharmaceuticals		
PS	Claim 7; Page 18-20; 26pp; English.		
CC	The present sequence encodes an aminocyclase enzyme. The enzyme is		
CC	capable of hydrolysing N-acetyl-D-tryptophan at a substrate		
CC	concentration of 10 grams per litre, and exhibiting faster conversion		
CC	of R-N-acetyl-2-thienylalanine than of		
CC	R-N-acetyl-4-chlorophenylalanine. The enzyme can be used for the		
CC	preparation of a D-amino acid by conversion of a corresponding		
CC	D-N-acetylamino acid. It can also be used for resolving a racemic		
CC	mixture of N-acylamino acids and deprotecting optically-enriched		
CC	N-acyl amino acids. The D-amino acids are useful as intermediates in		
CC	the production of various pesticides, antibiotics and other		
CC	pharmaceuticals.		
SQ	Sequence 1473 BP: 227 A; 559 C; 481 G; 206 T; 0 other:		
Query Match	81.9%; Score 1440.6; DB 21; Length 1473;		
Best Local Similarity	99.0%; Pred. No. 2.9e-217;		
Matches 1449; Conservative	0; Mismatches 14; Indels 0; Gaps		
0;			
0y	31 TCATGTCGCAGTCCGATTCCAGGCCCTTGACGCGTGTGGGGGGGCACCCTCATC	90	
Db			
	4 TCATGCGCCGAATCCGATTCCAGGCCCTTGACGCGTGTGGGGGGGCACCCTCATC	63	
0y	GACGGCAGCAACACCCC	150	
Db			
	64 GACGGCAGCAACACCCC	123	
0y	151 GCCATGCGGATCTGTGGACGCGCGCGGACACACC	210	
Db			
	124 GCCATGCGGATCTGTGGACGCGCGCGGACACACC	183	

OY	211	GTCCGCGCCGGCTTCATCGACTGCGACACCCAGACAGACAACTACTCTGTACGGCTGCG	270
Db	184	GTCCGCGCCGGCTTCATCGACTGCGACACCCAGACAGAACTACTCTGTACGGCGCCG	243
OY	271	GACATGACGCCCAAGATCTGCGACAGGGCGTCCACACGGTGTACAGGGGAATTGGCGCATC	330
Db	244	GACATGACGCCCAAGATCTGCGACAGGGCGTCCACACGGTGTACAGGGGAATTGGCGCATC	303
OY	331	AGCCTGGGCGCGTGGCGCACGCCAAACCGCGCCCGCTTGAGACTCTGTGACGAAGGC	390
Db	304	AGCCTGGGCGCGTGGCGCACGCCAAACCGCGCCCGCTTGAGACTCTGTGAGGAAGGC	363
OY	391	GGCTCTTACCGTTTCGAGCGCTTCGCGGACTACTCTGGAAGCGTTGCGGGGACCGCGCGG	450
Db	364	GGTTCTGTAACCGCTTCGAGCGCTTCGCGGACTACTCTGGAAGCGTTGCGGGGACCGCGG	423
OY	451	GCCCTCAACGCCGCTGTATGTGGGCCATTCAACGCTGCGCGCGCGGTATGCGCGGAC	510
Db	424	GCCCTCAACGCCGCTGTATGTGGGCCATTCAACGCTGCGCGCGCGGTATGCGCGGAC	483
OY	511	TTTGAGAGCGCGCGCACCGACGAGAGAAATCGCGGCGCATCGGGGACTGGCGGAGAAACC	570
Db	484	TTTGAGAGCGCGCGCACCGACGAGAGAAATCGCGGCGCATCGGGGACTGGCGGAGAAACC	543
OY	571	ATGGCCAGCGCGGCATCGGCATTTCGACCGGCGGCGCTTACCGCGCGCGCGCGCGCC	630
Db	544	ATGGCCAGCGCGGCATCGGCATTTCGACCGGCGGCGCTTACCGCGCGCGCGCGCGCC	603
OY	631	ACCCACCGAAGAGATCATCGAGGTGTGCGCGCGCTAGCGCGCATATGGCGCATCTAGCC	690
Db	604	ACCCACCGAAGAGATCATCGAGGTGTGCGCGCGCTAGCGCGCATATGGCGCATCTAGCC	663
OY	691	ACCCACATGCGCGCAGAGAGCGCAGACATGTCGGCGCGCTGTAGAGAAACTTCGCGCATC	750
Db	664	ACCCACATGCGCGCAGAGAGCGCAGACATGTCGGCGCGCTGTAGAGAAACTTCGCGCATC	723
OY	751	GGCGCGAGCTGGAAGTGTCGGGTGTGATCTGCGACACAGAGTCATGGCGCACGCCAAT	810
Db	724	GGCGCGAGCTGGAAGTGTCGGGTGTGATCTGCGACACAGAGTCATGGCGCACGCCAAT	783
OY	811	TTCCGCGCGCTGGCGCGAGAGCGTGGCGGTATGAGAGCGCCATGGCGCGCAGAGAGTC	870
Db	784	TTCCGCGCGCTGGCGCGAGAGCGTGGCGGTATGAGAGCGCCATGGCGCGCAGAGAGTC	843
OY	871	TGCGTGAAGCGGTATTCCTTACGTGGCGGCTCCACCATTCTCAAGCAGAGCGGTGCTG	930
Db	844	TGCGTGAAGCGGTATTCCTTACGTGGCGGCTCCACCATTCTCAAGCAGAGCGGTGCTG	903
OY	931	CTGGCGCGGAGCGCACATCATCATCACTGGTGAAGCCTTCCCGAACTGAGGGGCGCGAC	990
Db	904	CTGGCGCGGAGCGCACATCATCATCACTGGTGAAGCCTTCCCGAACTGAGGGGCGCGAC	963
OY	991	CTGATGAAGTGGCGGCGCGAGCGCGGCAATCCAAATGACAGTGTGTCGGAGTGTAG	1050
Db	964	CTGATGAAGTGGCGGCGCGAGCGCGGCAATCCAAATGACAGTGTGTCGGAGTGTAG	1023
OY	1051	CCGGCGCGGCGCATCTACTTCATGATGAGAGAAACCGGACGTGACGCAATCTTGCGCTTC	1110
Db	1024	CCGGCGCGGCGCATCTACTTCATGATGAGAGAAACCGGACGTGACGCAATCTTGCGCTTC	1083
OY	1111	GGCCGAGACATGATCGGCTCGAGCGGCTGTGCGGACAGAGAGCGCGCGATTCGCGGCTG	1170
Db	1084	GGCCGAGACATGATCGGCTCGAGCGGCTGTGCGGACAGAGAGCGCGCGATTCGCGGCTG	1143
OY	1171	TGGGCGACCTTCCGCGGCGGTCTGGGCGCACTAGCGCGAGACTGGGCTGTTCGCCGTG	1230
Db	1144	TGGGCGACCTTCCGCGGCGGTCTGGGCGCACTAGCGCGAGACTGGGCTGTTCGCCGTG	1203
OY	1231	GAGAGCGCGGTATGAGATGAGACCGGCTGACCGCGCGCGCTTGGGCGTGTGGCGGCGCG	1290
Db	1204	GAGAGCGCGGTATGAGATGAGACCGGCTGACCGCGCGCGCTTGGGCGTGTGGCGGCGCG	1263
OY	1291	GGGAGGCTGACAGGCGCGGCTTCTCGCGGACCTGTGTGATCCCGGCGACCGGTGCGC	1350

Db	1264	GGGAGAGCTGCAAGGCGGCTACTTTCGCCCACTGTGGGTCTTCGACCCGGCAGCGGTGGCC	1323
QY	1351	GATACCGCCACCTTCGAGAACACCTTACGAGGCGCGCGCGGCATCATTCCGTATACGTC	1410
Db	1324	GATACCGCCACCTTCGAGAACACCTTACGAGGCGCGCGGCATCATTCCGTATACGTC	1383
QY	1411	AAGGCGCGCGGCTGTGGCAAGAGAGGCGCTTCACCGGCGACATCGCGCGCGGCGTC	1470
Db	1384	AAGCGCGCGCGGCTGTGGCAAGAGAGGCGCTTCACCGGCGACATCGCGCGCGGCTTC	1443
QY	1471	GCACGCGACGCGCGCTTCGAGCCCG 1493	
Db	1444	GCACGCGACGCGCGCTTCGAGCTCG 1466	
RESULT 3			
ID	AAK53491/c		
AC	AAK53491 standard; DNA: 114955 BP.		
XX	AAK53491;		
DT	05-JUL-1999 (first entry)		
XX			
De	Human adenosine A1 receptor antisense oligonucleotide fragment.		
KM	Antisense oligonucleotide; multiple target; antisense treatment;		
KM	impaired respiration; inflammation; lung disease;		
KM	pulmonary vasoconstriction; inflammation; allergic rhinitis;		
KM	acute asthma; allergy; asthma; impeded respiration;		
KM	respiratory distress syndrome; pain; cystic fibrosis;		
KM	pulmonary hypertension; pulmonary vasoconstriction; emphysema;		
KM	chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;		
KM	colon cancer; breast cancer; lung cancer; pancreatic cancer;		
KM	hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;		
KM	prostate cancer; ss.		
OS	Synthetic.		
XX			
PN	WO9913886-A1.		
XX			
PD	25-MAR-1999.		
XX			
PP	17-SEP-1998; 98WO-US19419.		
XX			
PR	09-JUN-1998; 98US-0093972.		
PR	17-SEP-1997; 97US-0059160.		
XX			
PA	(UYEC-) UNIV EAST CAROLINA.		
PI			
NYce JW;			
XX			
DR	WPI: 1999-229400/19.		
XX			
PT	New antisense oligonucleotides used in treatment of, e.g. pulmonary		
PT	vasoconstriction		
XX			
PS	Disclosure: Page 37; 120pp; English.		
XX			
CC	The specification describes antisense oligonucleotides (AAK52869-X55271)		
CC	directed against at least 2 mRNAs selected from target genes, coding and		
CC	non-coding regions of RNAs corresponding to target genes, gene		
CC	initiation codons, genomic flanking regions, intron-exon borders, the		
CC	5'-end, the 3'-end and the juxta-section between coding and non-coding		
CC	regions and all segments of RNAs encoding proteins associated with one		
CC	or more diseases, conditions or mixtures. The antisense oligonucleotides		
CC	may be derived from sequences AAK5272-74. These multiple target		
CC	oligonucleotides (specifically AAK5180-271) can be used for the		
CC	antisense treatment of diseases and conditions. Typical diseases and		
CC	conditions are those associated with impaired respiration and		
CC	inflammation, including lung diseases, pulmonary vasoconstriction,		
CC	inflammation, allergic rhinitis, acute asthma, allergic, asthma, impeded		
CC	respiration, respiratory distress syndrome, pain, cystic fibrosis,		

CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer, CC pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma, CC hepatic metastases, as well as all types of cancers which may metastasize or have metastasized to the lungs including breast and prostate cancer.

SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Query Match	Score	DB	Length
5.98;	103;	20;	114955;

Matches 501; Conservative 155; Mismatches 952; Indels 9; Gaps 4;

Qy 110 GGGGGGGCGGACCTGGGCGTGGCGGGGCGAACCGCATCGCGCGATGTGCG 169  
||||||| : || | | | | | : : || | | | |  
Db 105211 GCGGGCGCGCCNNNNNNNSCGGCGCGCGGCGCGCCCVNNNNNSCGGCGCGCG 105152

170 ACG--CCGCCGCCACACCCGGGTCGACGTGTCGGGCTGGTGGTCGCCGCCGCTTCAT 227

Db 105151 GCGGCGCCCGVGNHNNNSCGGCCGCGCGCGCGCGCCCGVGNHNNNSCGGCCG 10509

QY 228 CGACTCGCACACCCACGACGACAACTACCTGCTCAGGCGTCGCGACATGACGCCCAAGAT 287

Db 105091 CCGGGCGCGCCCVGGC<sup>1</sup>NNHNNNSCGGCCCGGCCGGCGGCCCGCCVGGC<sup>2</sup>NNHNNNSC 105

288 CTCGAGGGCGTCACCAAGGTGGTCACGGGCAATTGGGCATCAGCCTGGCGCGCTGGC 347

Db 105031 GGGCCGGCGGGCGCCCGCCVGGCCVNNHNNNSCGGCCGGCGCGCCCGCCVGGC 104972

348 GCACGCCAACCCGCCGCCCTGGACCTGCTGGACGAAGCGGCTCTTACC GTT CGA 407

Db 104971 CVGNHNNSCGGCCGGCGCCCGCCVGGCCVGCNNHNNSCGGCCGGCC 104

408 GCGCTTGGCCGACTACCTGGACGCGTTGCGGGCCACGCCGGCGCGTCAACGCCGCGCTG 467

Db 104911 GGGGGCCVGGCCVGGGNNHNNNSGGCCVGGGNNHNNNSVGGCCVGGGNNHNNNSCV 104

468 TATGGTGGCCATTCAACGCTGCGCGCGGGTCATGCCGACTTGCAGCGCGCGCCAC 527

Db 104851 GGCCTGGGNNNNNSCCVGGCCVGGGNNNNNSCCCVGGCCVGGGNNNNNSGCCCV 104

528 CGACGAGGAATCGCGGCATGCGGGACCTGGCCGAGGAGCCATGGCCAGCGCGCCAT 587

Db 104791 GGGCVGGGNNNNNSGGCCVGGCCVGGGNNNNNSGGCCVGGCCVGGGNNNNN 104

588 CCGCATTTTCGACCGGCGCCTTCTACCCGCGCCCGCCGCGCCACCAACGAGAGATCAT 647

Db 104731 SCGCGCCCVGGCCVGGGNNNNNSGCGGCCCVGGCCVGGGNNNNNSGCGGCCCV 104

648 CGAGGTGTGCGCGCGCTGAGCGCGGCATGGCGGCATCTACGCCACCCACATGGCGGACGA 707

Db\_104671 GGCVCGGNNHNNNSCGCGCCGCCC VGCCVCGGNNHNNNSGCGCGGCCCVG--GC 104

708 AGCGGAGCATCGTGGCCGCGCTGGAGGAACCTCCGCATCGGCCGCGAGCTGGACGT 767

Db 104613 C VCGGNNHNNNSGGCGGCGGCCCVGGCCVGGGNNHNNNSCGGCGGCGGCCCVGGCC 104

QY 768 GCCGGTGTGATCTCGCACCAAGTTCATGGCCAGCCCAATTTCGGCCGCTCGCGCGA 82

D6 104553 VCGGNNHNNNSCCGGCGCGCGCCVGGCVCGGNNHNNNSGCCGGCGCGCCV6 104

828 GACGCTGCCCGTGAACGAGCCCGCCATGGCGCCCGAGGACGTCTCGCTGGACCGCTATCC 88+

[illegible]

888 CTACGTGGCCGGCTCCACCAAGCTCAAGCAGGACCGGCTGCTGGCCGACCAAT 944

DB 104433 CGCCVGGCCVGC GGNHNNNSC ---CGCCCGCGCGCGCCVGGCCVCGGNNHNNNS 104

340 CALCACCGGCAAGCCCTCCCCGAACGGAGCGGCGCGAAGCAGGC 100

QY | | | | | : | | | | | : | | | | | :

DB 104378 CCGGCGGGGCGGCGCCVGGC VGGNNNHNNSGCGCGGCGGCGGCGCCVGGC 104

1





Dp	20298	TCACACACCCCTCAGACACACACACCCCGGCGCGCCCGCCGTCGTCGCGAGCCGCGACGCGGACGACGACACCG	20357
Dp	722	TGCGCCCGCCCTGGAGGAGAAACCTTCCGCATTCGCGCCCGCGAGCTGAGAGTGCCTGGTGTATCT	781
Dp	20358	CCGCGCTCTCTCGCCCTCGCCCAACAGTGGCTCGCCGAGACCCGTTGCGCGAGCGCCGCG	20417
Dp	782	CGCAGCACAAGGTGATGAGGCGCAGGCCAAATTTTCGCGCCGCTCGCGGAGACCTCTCCGCTGA	841
Dp	20418	TGGTCTCTGTCACACCGCGGCGCGACACGAGAGGACGAGACCCGCGCGCGCGCGCGCGG	20477
Dp	842	TCGAGGCGCCGATGCGCGCGCCAGAGAGCTCTGCTGAGAGCGTATCCCTAGTGGCCGGCT	901
Dp	20478	GCCCTGATTCGCGACCGCCGCCACCGAGAAACCCCGGCGCTTTGCGCTCTCTGACCTTCGCGC	20537
Dp	902	CCACCATGTGTCAGAGCAGAGACGCGTGTGCTGCGCGGAGCGACCATCATCTACCTGCTGGA	961
Dp	20538	CCGACACACCGCGCGCGCGCGACCCCGAGACCTCTGC-----CACGCGCCCTGGCCCGCCAGCC	20591
Dp	962	AGCCCTTCCCGGAACTGAGCGGGGCGCGACCTGGATGAATGCGCGGCGCGCGCAAAAT	1021
Dp	20592	ACGACAGGCGCCGACCTCGCGCTCGCCGCGACCGAGTCAAGCGCGCGCCGCTGGCCCTG	20651
Dp	1022	CCAAATGACGAGTGGGTGCGCCGAGTGCAGCGCGCGCGCGCCATTAATCTATGATGAGAG	1081
Dp	20652	TCCCGCTCTGCGACCGGACCCACACCACTTGGAAACCGGAGCGACGCTCTATACCGGCG	20711
Dp	1082	AACCCGACGTGACGCGCATCTGCGGCTTTCGCGCGCGACCATGATGCGCTCGACGAGCCTGC	1141
Dp	20712	GCAACCGCGGCGCTGGG---CGCGGTCTCTGCGCCCGCACATCTGTGTCCACCAACGCGGCTCC	20768
Dp	1142	CGACGACGAGGCGCCCGGATTCGCGCGCTGTGGGGGACCTTCCGCGGGGTCTGCGGGGCACT	1201
Dp	20769	GCCACCTGTCTGTCTGCGACACCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGA	20828
Dp	1202	ATGCGCGCGACACTGGGCGCTTTCGCGCTGAGAGACGCGCGGTATGGAAGATAGCGGCTGA	1261
Dp	20829	CCGAACTCAGACCGGGGCTCGGCGCGCACCGTCTCAATGCGCGCTGCGAGGTGCGGACCGG	20888
Dp	1262	CGCGCGCGCGCGCTTCGCGCGTGGCGGGCGCGCGCGACACTGACGACCGGGTACTTGGCGGACC	1321
Dp	20889	CCG-----CTCGCGCGACCTGTGTGGGACCGTTCGCGCGCGCGCGCGCGCGCGCTCACCG	20939
Dp	1322	TGGTGTGTTTCGACCCCGCGACGATGAGCGGAGATACCGCACCTTTCGAGACACCTTACCGAGC	1381
Dp	20940	TCTGTCTTCACACCGCGCGGCGTCTGAGAGAGCGGTCTCTGGCTCCCTTCACCCGCGACG	20999
Dp	1382	GGCGCGCGCGGATTCATTCCTGTCTACGTACGTACAGCGCGCGCGGCTGTGGCGAGAGCGGCT	1441
Dp	21000	GCTGTGACACCGTCTCGCGCGGCCCAAGGCGAGACGCGCTGTGCACCTGCAGAGGCGACCC	21058
Dp	1442	TCACCGGCGCAGATGCGCGCGCGCGGTCTGCGACGACGCGCGCGCTGTGAGCGCGGCGCGAG	1501
Dp	21060	GCCACCTCTGACCTGAGAGGCTTCTCTCTTCTGTCTCGTGGCGCGCGCACCTCTGGGACGC	21119
Dp	1502	CTTTAC-----AATCGGCGGTGAACGCGGCGGCGTGCAGCGCCCTTCCCAACCTCTGAGCG	1555
Dp	21120	CCGAGCAGGCGCACTACGCGCGCGGCAACGACCTTCTGTGAAGCGCTCTGCGCGCGCGCGCG	21179
Dp	1556	CAAAACGCTCATAGGCGCGGCTCTCGCGTCCGCAATAGCGCGCGCACCCGAGATATGTGGGA	1615
Dp	21180	CCGCGACCGGCGCTGCGCGCGCACCTCTCTGCTTGGGCGCGCGTGGACCCAGAGGCTGTGGCA	21233
Dp	1616	AGGAAGTGAATGGCGCGCGCTGCGCGCGCGAGCGACGACGCGCG---GAAATAGACCTTCG	1672
Dp	21240	TGACAAACCACTCTGTCTGACACTGTGAGCTGTGAGCTGACGCGATGCGCGGATGCGCGCGCG	21299
Dp	1673	AAGACCTGTGCGAGGCGCGACGCGATCTCGCGGTCTGACACCTGTGC	1716
Dp	21300	TGACCTCTGGAACGAGGCGACCGCGCTTCTTGACGCGCGCGCTGTGGCC	21343

XX	AA017186	standard; DNA: 125401 BP.
XX	AA017186	
AC	AA017186;	
XX		
DT	29-NOV-2001	(first entry)
XX		
DE	Streptomyces noursei	nystatin PKS gene cluster DNA.
XX		
KM	polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;	
XX	antifungal; antibiotic; ds.	
XX		
OS	Streptomyces noursei.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	6337..34771
FT		/*tag- a
FT	CDS	/product- "Nysi complete protein"
FT		34792..51099
FT		/*tag- b
FT	CDS	/product- "NysJ protein"
FT		51155..57355
FT		/*tag- c
FT	CDS	/product- "Nysk protein"
FT		57503..58687
FT		/*tag- d
FT	CDS	/product- "Nysl protein"
FT		complement (58786..58980)
FT		/*tag- e
FT	CDS	/product- "Nysm protein"
FT		/note- "CDS does not include start codon"
FT	CDS	complement (59045..60241)
FT		/*tag- f
FT		/product- "NysN protein"
FT		/note- "CDS does not include start codon"
FT	CDS	complement (60238..61296)
FT		/*tag- g
FT		/product- "NysD2 complete protein"
FT	CDS	120628..121308
FT		/*tag- h
FT		/product- "NysR4 (long) protein"
XX		
PM	W0200159126-A2.	
XX		
PD	16-AUG-2001.	
XX		
PF	08-FEB-2001; 2001WO-CB00509.	
XX		
PR	08-FEB-2000; 2000GB-0002840.	
PR	10-APR-2000; 2000GB-0008786.	
PR	14-APR-2000; 2000GB-0009387.	
XX		
PA	(UNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.	
PA	(SMTF) SINTEF STIFTELSEN IND TEK FORSK.	
PA	(ALPH-) ALPHARMA AS.	
PA	(SINV-) SINVENT AS.	
PA	(DZIE/) DZIELEWSKA H.	
PA	(ZOTC/) ZOTCHEV S B.	
PA	(SEKU/) SEKUROVA O N.	
PA	(FJAE/) FJAEVRIK E.	
PA	(BRAU/) BRAUTASET T.	
PA	(STRO/) STROM A R.	
XX		
PI	Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;	
PI	Valla S, Ellingsen TE, Sletta H, Gulliksen O;	
XX		
XX	WPI: 2001-557614/62.	
DR	P-ESDB: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148	
DR	AAE10149, AAE10150.	
XX		
XX	New nystatin polyketide synthase polynucleotides and polypeptides,	
XX	useful as antibiotics and antifungals -	



PS Claim 1: Page 188-254: 266bp; English.

CC The present invention relates to the cloning and sequencing of the gene  
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme  
 CC involved in the biosynthesis of the macrolide antibiotic nystatin.  
 CC The nystatin PKS is useful as antifungal antibiotics. The present  
 CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.

XX Sequence 125401 BP: 15664 A: 49692 C: 42871 G: 17174 T: 0 other:

Query Match 5.6%; Score 99.2; DB 22; Length 125401;  
 Best Local Similarity 43.5%; Pred. No. 1.7e-07;  
 Matches 724; Conservative 0; Mismatches 913; Indels 27; Gaps 5;

62 ACCGTGCTCTGCGGGGCGGACCTCATGCGGAGCAACACCCGGGGCGGGCGG 121  
 79959 ACGAGCCCTGCGGCGCACTGCGCTTCAGTACGCGCCCGCTTCCAGGGCTGCGGCGG 80018  
 122 ACGTGGCGTGGCGGGGACCGCATCGCGGCATCGGCGCATGTGTGGAGCGCGCGCGC 181  
 80019 CTTGGCGCGCGGACACGACGAGTACGCGGAGTGGCCCTGCGCGGAGCGCGGACCG 80078  
 182 ACACCGGGTGCAGTGTGCGGCTGTGCTGCGCGCGCTTTCATGACTGCGACACCC 241  
 80079 ACCCGCGCGCTTGGAGTGCAGCGCGCGCTGCTGAGCGCGGACAAACAGCGCGCGCT 80138  
 242 ACGAGACACTACCTGCTCAGGCGTGGCGGACATGAGCGCGGCAAGATCTGCGAGGGGTCA 301  
 80139 ACGCGGCTGCGGCGCATCGCGCGCGCGCGCTGCGCTTGGCTGGGAAGGCTCTCGC 80198  
 302 CCAGCGGTGTCAGGGGCAATTGCGGCATCAAGCTGCGCGCGCTGCGCGACGCCACCGCG 361  
 80199 TCG 80258  
 362 CCG 421  
 80259 TCACCATGCGCTTACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 80318  
 422 ACCTGAGCGCGTGTGCGGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 481  
 80319 CCGCGGAGTGTCCCGCT 80378  
 482 CAACGCTGCG 541  
 80379 TCCAGCTGCAAGTGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80438  
 542 CGGCGATGCGGCG 601  
 80439 TCTTCG 80498  
 602 GCGCGCTTACCG 661  
 80499 CCG 80558  
 662 CGCTGAGCG 721  
 80559 TCACCAACCTTCACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80618  
 722 TGGCG 781  
 80619 CCG 80678  
 782 CGCACCAAGTCAATGCG 841  
 80679 TGGCTCTGTCACCG 80738  
 842 TCGAGCG 901  
 80739 GCGTGAATCGGCG 80798  
 902 CCACCATGCTCAACGACGACGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 961  
 80799 CCGAGACG 80852

962 AGCCCTTCCCGAAGTACGAGCGGCGGACCTGAGTACGCTGCGCGCGCGCGCGCGCGCGCAAT 1021  
 80853 ACGAGAGCG 80912  
 1022 CCAAGTACGACGTGTGCG 1081  
 80913 TCCCGCTGCG 80972  
 1082 AACCGAGCG 1141  
 80973 GCACCG 81029  
 1142 CGCAGAGAGCG 1201  
 81030 GCGACGTGCTGCTGCG 81089  
 1202 ATGCG 81149  
 81090 CCGAATCAGCG 81199  
 1262 CCG 1321  
 81150 CCGC-----CGTGGCGAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81200  
 1322 TGTGCTGTGACCG 1381  
 81201 TGTGCTGTGACCG 81260  
 1382 GCG 1441  
 81261 GCGTGAACCG 81320  
 1442 TCACG 1501  
 81321 GCACCTGACG 81380  
 1502 CCGTAC-----ATCCCG 1555  
 81381 CGGAGAGCG 81440  
 1556 CAACCGTACATGCG 1615  
 81441 CCGCGCACCG 81500  
 1616 AGGAGTGTGCG 1672  
 81501 TGACAGCAGCG 1736  
 1673 AAGACGCTGCG 1796  
 81561 TGACCTGGAACAGGCG 81604

# RESULT 7

AAQ73500/c  
 ID AAQ73500 standard; DNA: 8438 BP..

XX AAQ73500:

XX 15-MAY-1995 (first entry)

XX DNA encoding Pseudorabies virus large latency transcript.

XX Pseudorabies virus; PRV; LNV; large latency transcript;

XX attenuated virus; vaccine; early protein 0; Ep0; HSV-1, ICP0;

XX protecting animals; deletion mutants; swine; ds.

XX Pseudorabies virus.

XX Key Location/Qualifiers

FT misc\_feature 1..7013

FT tag- a







OY	559	GC	CAGAGAA	AGCCAT	GGCAGC	GGGCGCCAT	TTCGAC	GGGCGCTT	TACCCGCC	618
Db	24029	ACC	CTCGGG	CTGACGG	GTCCAC	CAACCA	CGACCG	CGCGCGCC	CGGATACCGAA	24088
OY	619	GG	CGCCCG	CGCCAC	CACCGAG	AGATCAT	TCAGAG	GTGCGCGCGCT	ATGAGCGCAT	678
Db	24089	GCC	CGCCGCG	CAACAG	GGCCCG	CTTAA	GCGGTGT	GCTGTCT	CTCTGCGCT	24148
OY	679	GG	CATCTA	ACGGC	ACCA	CCATGCG	AGCAAG	CGACGAC	ATCTG	738
Db	24149	GAC	CGCGGC	CAACCC	CGGTGG	CGCGCC	CCCTCA	CCCTCA	CAACAC	24208
OY	739	AC	CTTCCG	CATCG	CGCGG	AGCTG	AGC	CGGTGG	TGATCTG	798
Db	24209	CT	CGCGCA	CGCGCG	CATGAG	CGCGCG	CTG	TGGAAC	GTACAC	24268
OY	799	GG	CCAGC	CCCAAT	TTCGG	CGCGCT	CGCGCA	GAAGCG	CTGATCG	858
Db	24269	GG	CGCGCG	CGCAAG	GTATAC	CGCGCC	CA	-----	CA	24319
OY	859	CG	CAAG	CACT	CTCGCT	GACCG	GTATCT	CACT	AGTGC	918
Db	24320	GG	CGCGCG	CGCTCG	CCCTG	GAAC	TGCGCG	CGCTT	CGCGCA	24379
OY	919	GAC	CGCGT	GTGCT	GGCGG	AGCA	CAATCA	CTG	TGCAAG	978
Db	24380	ACC	CTTG	GAAC	GGCC	CAAGCC	CGCGCG	GTGCG	GTGCGGT	24439
OY	979	AG	CGGCG	CGCA	CTGAT	GAAAG	TGCGCG	CGCA	CGCGCA	1038
Db	24440	GAC	CGG	GTGG	CGCTG	CGCGCG	CTTCT	CTTCT	CGCGCG	24496
OY	1039	CC	CGA	CTG	CA	CGCGCG	CGCAT	CTTCA	TGATG	1098
Db	24497	G	CGCGCG	CGCA	ACCGCG	CGCGCA	CCCGCT	CGCA	CGCGCG	24556
OY	1099	AT	CTG	GGG	GTG	CGCGCA	CTGAT	GCGCT	CGCGCA	1158
Db	24557	G	CGCGCA	CGCGCG	CACTG	CGCGCA	CGCGCG	CGCGCG	CGCGCG	24616
OY	1159	CA	TCCG	CGCGCT	-----	GGG	CACT	TTC	CGCGCG	1215
Db	24617	CA	CTG	CTCT	CA	CGCGCG	CGCGCG	CGCGCG	CGCGCG	24676
OY	1216	G	CGCT	GTTC	CGCGCT	GA	CA	CGCGCG	GTATG	1275
Db	24677	G	AAC	TGGA	GAAC	CTGG	CGCGCG	CGCGCG	CGCGCG	24736
OY	1276	G	AC	CTG	CGCGCG	CGCGCA	CTG	GA	CGCGCG	1335
Db	24737	G	CGCGT	CGCGCG	CTCT	CGCGCA	CTG	CGCGCA	CGCGCG	24796
OY	1336	CG	CGCC	CA	CGGT	AGAT	CGCC	CACT	-----	1392
Db	24797	ACC	CGCGCG	GTGT	GTGAG	AGCA	CGCTG	GTGAG	CGCGCG	24856
OY	1393	AT	CCAT	TCCG	GTAC	GTCA	AGGCG	CGCGGT	GTGCA	1452
Db	24857	G	TGCT	CGCGCA	AGCA	CGCTCG	CGCGCA	CTG	CA	24916
OY	1453	CA	TG	CGCGCG	GTGCT	CTG	CA	CGCGCG	GTGAG	1512
Db	24917	CT	CGCGCG	TTGT	GTGCT	TTCT	CTCA	CGCGCG	GTCTG	24974
OY	1513	G	CGGT	GA	AGGCG	CGCGCG	CGCGCG	CGCGCG	CGCGCG	1563
Db	24974	G	GA	CTAC	CGCGCG	CAAC	CGCGCA	CGCGCG	CGCGCG	25024

```

RESULT 10
AAL40781
ID      AAL40781 standard; DNA; 88421 BP.
XX

```

AC	AAL40781;
XX	03-OCT-2002 (first entry)
DT	
DE	8842Int genomic DNA of ramoplanin producing Actinoplanes sp.
XX	
KW	Ramoplanin; ramoplanin biosynthetic pathway polypeptide;; anti-
KM	biosynthesis gene cluster; bioengineering; peptide synthetase
KW	adenylation domain; hydroxyphenylglycine; HPG; antibiotic pe-
KV	chlorinate; lippseptide; gene; ds.
XX	
OS	Actinoplanes sp.
XX	
FH	Location/Qualifiers
FT	CDS
FT	/tag= a
FT	/product= "Protein of ORF 1"
FT	3118..4032
FT	/tag= b
FT	/product= "Protein of ORF 2"
FT	4038..5048
FT	/tag= c
FT	/product= "Protein of ORF 3"
FT	complement (6665..5814)
FT	/tag= d
FT	/product= "Protein of ORF 4"
FT	complement (7703..6693)
FT	/tag= e
FT	/product= "Protein of ORF 5"
FT	complement (9464..8130)
FT	/tag= f
FT	/product= "Protein of ORF 6"
FT	9691..10761
FT	/tag= g
FT	/product= "Protein of ORF 7"
FT	complement (12751..10829)
FT	/tag= h
FT	/product= "Protein of ORF 8"
FT	complement (13617..12802)
FT	/tag= i
FT	/product= "Protein of ORF 9"
FT	complement (15203..13614)
FT	/tag= j
FT	/product= "Protein of ORF 10"
FT	15591..15863
FT	/tag= k
FT	/product= "Protein of ORF 11"
FT	15880..19035
FT	/tag= l
FT	/product= "Protein of ORF 12"
FT	19032..39713
FT	/tag= m
FT	/product= "Protein of ORF 13"
FT	39713..65800
FT	/tag= n
FT	/product= "Protein of ORF 14"
FT	65826..66530
FT	/tag= o
FT	/product= "Protein of ORF 15"
FT	66546..67370
FT	/tag= p
FT	/product= "Protein of ORF 16"
FT	67384..70059
FT	/tag= q
FT	/product= "Protein of ORF 17"
FT	70099..70662
FT	/tag= r
FT	/product= "Protein of ORF 18"
FT	70659..71906
FT	/tag= s
FT	/product= "Protein of ORF 19"
FT	complement (73439..71964)
FT	/tag= t
FT	CDS

```

/product= "Protein of ORF 19"
complement (73439..71964)
/tag= t

```

FT	/product- "Protein of ORF 20"
FT	complement (74216..73563)
FT	/tag- u
FT	/product- "Protein of ORF 21"
FT	complement (75424..74213)
FT	/tag- v
FT	/product- "Protein of ORF 22"
FT	75535..76464
FT	/tag- w
FT	/product- "Protein of ORF 23"
FT	complement (78110..76449)
FT	/tag- x
FT	/product- "Protein of ORF 24"
FT	complement (79864..78107)
FT	/tag- y
FT	/product- "Protein of ORF 25"
FT	complement (81624..79861)
FT	/tag- z
FT	/product- "Protein of ORF 26"
FT	complement (81909..81682)
FT	/tag- aa
FT	/product- "Protein of ORF 27"
FT	complement (82346..82062)
FT	/tag- ab
FT	/product- "Protein of ORF 28"
FT	82587..84446
FT	/tag- ac
FT	/product- "Protein of ORF 29"
FT	84481..85548
FT	/tag- ad
FT	/product- "Protein of ORF 30"
FT	85556..86845
FT	/tag- ae
FT	/product- "Protein of ORF 31"
FT	87372..86803
FT	/tag- af
FT	/product- "Protein of ORF 32"
FT	87494..86420
FT	/tag- ag
FT	/product- "Protein of ORF 33"
FT	
XX	MO200231155-A2.
PN	
PD	18-APR-2002.
XX	
XX	15-OCT-2001; 2001WO-CM01462.
PR	13-OCT-2000; 2000US-239924P.
PR	12-APR-2001; 2001US-283286P.
PR	24-JUL-2001; 2001US-0910613.
PA	(ECOP-) ECOPIA BIOSCIENCES INC.
PI	Farnet CM, Zazopoulos E, Staffa A;
PJ	
XX	WPI: 2002-435445/46.
DR	P-PSDB: AAO221166, AAO22117, AAO22148, AAO22149, AAO22150, AAO22151,
DR	AAO22152, AAO22153, AAO22154, AAO22155, AAO22156, AAO22157, AAO22158,
DR	AAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165,
DR	AAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,
DR	AAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.
XX	
PT	Novel isolated ramoplanin biosynthetic pathway polypeptide useful for
PT	chemically modifying biological molecule that is a substrate for a
PT	polypeptide encoded by a ramoplanin biosynthesis gene cluster -
XX	
PS	Disclosure: Page 87-135; 212pp; English.
XX	

The invention relates to an isolated ramoplanin biosynthetic pathway polypeptide selected from a polypeptide of open reading frames (ORF) 1-32. The isolated polypeptides are useful for chemically modifying a biological molecule that is a substrate for a polypeptide encoded by a ramoplanin biosynthesis gene cluster, by contacting the biological ramoplanin biosynthesis gene cluster.

molecule with the isolated polypeptide, where the polypeptide chemically modifies the biological molecule. The method comprises contacting the biological molecule with at least two different polypeptides encoded by ramoplanin ORFs 1-31. The polypeptides are useful for directing the biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated gene cluster comprising the ORFs is useful as a substrate for bioengineering of antibiotic structures. An isolated polypeptide or its encoding nucleic acid sequence is useful for generating derivatives of ramoplanin, for improving production or for producing variants of other antibiotics of the peptide class. The isolated polypeptides are useful for synthesis of ramoplanin *in vivo* or *in vitro*, as an adenylation domain in conjunction with other peptide synthetase modules and allowing the incorporation of Thr into a peptide antibiotic precursor, for modifying fatty acid structure and/or enhancing fatty acid incorporation into the peptide antibiotic structure, for production of an hydroxyphenylglycine (Hpg)-containing peptide antibiotic, for enhancing secretion of ramoplanin or its variants and derivatives, for enhancing uptake of precursors for ramoplanin biosynthesis, for enhancing production of ramoplanin products or its variants or derivatives, to chlorinate Hpg of a peptide antibiotic precursor, and for designing specific nucleotide probes and primers for identifying and isolating putative lipdipeptide-producing microorganisms. This polynucleotide sequence represents the 8842nt genomic DNA of a ramoplanin producing *Actinoplanes* sp. microorganism of the invention.

Q	Sequence	88421 BP;	10639 A;	32391 C;	32663 G;	12728 T;	0 other;
Query Match		4.98;	Score 86;	DB 24;	Length 88421;		
Best Local Similarity		43.18;	Pred. No. 2e-05;				
Matches 605;	Conservative	0;	Mismatches 775;	Indels 24;	Gaps 3;		

[illegible]



```
Db 46830 TCATGACCTGTCAGAGCCGCGCTGCGCTCTGCTGCTGCGGCTGCGGCGCGGACCGGACG 46889
Qy 689 CCACCCATGTCGCGAGAGGAGGACATCTGTCGCGCGCTGAGGAAACCTTCGCA 748
Db 46890 TCCGATCGGAGACGTCGTGCGCGCGCGCGGACCGGCGGCTGAGAACCTGCTGCGCT 46949
Qy 749 TCGGCGCGGAGCTGAGCTGCGGCTGCTGATCTGCAACACACAGGTATGAGCCAGCCCA 808
Db 46950 TCTCTGTCACACACCTGCTGCTGCGCACCGACCTGACCGGCGGCGCGCTGACCGAGC 47009
Qy 809 ATTTCGCGCGCTGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 868
Db 47010 TGTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 47069
Qy 869 TCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 928
Db 47070 AGAAGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 47123
Qy 929 TGTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 988
Db 47124 TCCAGGTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 47183
Qy 989 ACCGAGTGAAGTGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1048
Db 47184 CGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 47243
Qy 1049 AGCGGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1108
Db 47244 CCGAGACCTTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47303
Qy 1109 TCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1168
Db 47304 CCGACCTTTCGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 47363
Qy 1169 TGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1228
Db 47364 GCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 47411
Qy 1229 TGGAGACGCGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1288
Db 47412 TGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 47471
Qy 1289 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1348
Db 47472 CCGAGATCAGGTCGCGCGGCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 47531
Qy 1349 CCGATACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1408
Db 47532 CCGTACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 47591
Qy 1409 TCACGCGCGCGCGCTGCGCAAG 1432
Db 47592 TCGCGCGCTGCGCGCGCGCAG 47615

RESULT 11
AAS08693
ID AAS08693 standard; DNA; 109519 BP.
XX
AC AAS08693;
XX
DT 26-SEP-2001 (first entry)
XX
DE Micromonospora DNA encoding biosynthetic enzymes for Evernimycin.
XX
KW Evernimycin; antibiotic; bottle-neck gene; orthomycin;
KW fermentation; ds.
XX
OS Micromonospora carbonacea var. africana.
XX
FH Key Location/Qualifiers
FT CDS complement (132..1382)
FT /*tag- a
```

```

FT CDS      /*tag- ab
              22748..24172
FT          /*tag- ac
FT          /product- "EVRG"
FT RBS      22736..22740
FT          /*tag- ad
FT CDS      complement (24177..25223)
FT          /*tag- ae
FT          /product- "EVRH"
FT RBS      complement (25230..25233)
FT          /*tag- af
FT CDS      25550..26626
FT          /*tag- ag
FT          /product- "EVRT"
FT CDS      26685..30479
FT          /*tag- ah
FT          /product- "EVRJ"
FT RBS      26672..26676
FT          /*tag- ai
FT CDS      complement (30557..31876)
FT          /*tag- aj
FT          /product- "EVRK"
FT RBS      complement (31885..31888)
FT          /*tag- ak
FT CDS      complement (31941..32882)
FT          /*tag- al
FT          /product- "EVRL"
FT CDS      complement (33167..34405)
FT          /*tag- am
FT          /product- "EVRM"
FT RBS      complement (34414..34418)
FT          /*tag- an
FT CDS      complement (34449..35210)
FT          /*tag- ao
FT          /product- "EVRN"
FT RBS      complement (35219..35221)
FT          /*tag- ap
FT CDS      complement (35294..36238)
FT          /*tag- aq
FT          /product- "EVRQ"
FT CDS      complement (36235..36963)
FT          /*tag- ar
FT          /product- "EVRP"
FT CDS      complement (36998..38026)
FT          /*tag- as
FT          /product- "EVRQ"
FT CDS      complement (38072..38566)
FT          /*tag- at
FT          /product- "EVRP"
FT CDS      complement (38892..40163)
FT          /*tag- au
FT          /product- "EVRs"
FT CDS      complement (40216..40890)
FT          /*tag- av
FT          /product- "EVRT"
FT RBS      complement (40899..40902)
FT          /*tag- aw
FT CDS      complement (40887..41576)
FT          /*tag- ax
FT          /product- "EVRU"
FT CDS      complement (41679..42707)
FT          /*tag- ay
FT          /product- "EVRV"
FT RBS      complement (42714..42717)
FT          /*tag- az
FT CDS      complement (42810..43799)
FT          /*tag- ba
FT          /product- "EVRW"
FT RBS      complement (43807..43811)
FT          /*tag- bb
FT CDS      complement (43799..44866)
FT          /*tag- bc
FT          /product- "EVRX"

FT CDS      complement (45014..45760)
FT          /*tag- bd
FT          /product- "EVRy"
FT RBS      complement (45767..45770)
FT          /*tag- be
FT CDS      complement (45962..46714)
FT          /*tag- bf
FT          /product- "EVRz"
FT RBS      complement (45952..45956)
FT          /*tag- bg
FT CDS      complement (47156..49234)
FT          /*tag- bh
FT          /product- "EVRa"
FT CDS      51627..52715
FT          /*tag- bi
FT          /product- "EVRB"
FT RBS      51629..51622
FT          /*tag- bj
FT          52889..53557
FT CDS      /*tag- bk
FT          /product- "EVRc"
FT          53554..54207
FT          /*tag- bl
FT CDS      /product- "EVRd"
FT          complement (54362..55117)
FT          /*tag- bm
FT          /product- "EVRB"
FT RBS      complement (55125..55128)
FT          /*tag- bn
FT CDS      complement (55135..56094)
FT          /*tag- bo
FT          /product- "EVRc"
FT RBS      complement (56100..56103)
FT          /*tag- bp
FT CDS      complement (56184..56813)
FT          /*tag- bq
FT          /product- "EVRc2"
FT CDS      56961..58709

Query Match      4.88; Score 84.6; DB 22; Length 109519;
Best Local Similarity 45.18; Pred. No. 3.2e-05;
Matches 542; Conservative 0; Mismatches 629; Indels 30; Gaps 5;

QY 117 CGCGACCTGGGCGCTGGCGCGACCGCATCGCGCATCGCGCATCGCGCGC 176
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 94498 CGTGACCTGTTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94557

QY 177 CGCGACACCGCGGCGTGCAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCG 236
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 94558 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94617

QY 237 CACCCAGGACACACTACTGCTCAGCGCGTGCAGCATACGCCCAAGATTCG 236
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 94618 CGGCTTGCGCGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94677

QY 297 CGTCACACGCGTGTGTCACGCGCAATTCGCGCATCAGCGCGCGCGCGCGCA 356
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 94678 GGTACCGCGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94737

QY 357 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 404
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 94738 CTTGCGCGCGTGCAGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94797

QY 405 CGACCGCTTGCGCGACTGACGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 464
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 94798 CCGGCGCGGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94857

QY 465 CTGTATGCGCGCATTCACGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCG 524
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 94858 CGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 94917

QY 525 CACGACAGGAAATCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 584
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```



```

Db 395 CCGCCGCGGTACACCCCGACCTCGCGCGGTGCGGCTCCACCTCTGGGGCGCCCT 454
Oy 272 ACATGACGCGCCAAAGATTCGCGAGGGGTACACAGGTGTCAAGGGCAATTCGGCATCA 331
Db 455 GCGCCGCGGACGAGGTGCGGAAAGTGGCGGAGACACGCGCTGGCGGTACTTCTGACG 514
Oy 332 GCGTGGCGCGCGTGGCGGACGCAACCGCGCCCGCTGGACCTGCGTGGAGCAAGAGCG 391
Db 515 CCGGCGACGCGCTCGCGCTCGCGGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 574
Oy 352 GCTCTTACCGTTTCGAGCGCTTGGCGACTACCTGAGACCGCTTGGCGCGCGCGCG 451
Db 575 AGGTCTTACGCTTCCA---CGCCACCAAGCGGTCAACGCTTTCGAGGGCGCGCGCTG 631
Oy 452 CCGTAACGCGCGCTGTATGTTGGGCGCATTCACAGCTGCGCGCGCGCGCGCGCGACT 511
Db 632 TCACGAGACGCGCGCGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 691
Oy 512 TCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 571
Db 692 TCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 751
Oy 572 TGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 631
Db 752 TGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 811
Oy 632 CCACGAGAGAGATCATGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 691
Db 812 CCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 871
Oy 692 CCGACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 751
Db 872 ACAGCGCTCAACACACAGTACGTGATCTGTGAGATGAGAGAGCGCGCGCGCGCGCG 931
Oy 752 GCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 811
Db 932 ACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 991
Oy 812 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 862
Db 992 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1051
Oy 863 AGGACGCTCTCGCTGAGCGCGTATCCCTACGTGCGCGCGCGCGCGCGCGCGCGCG 922
Db 1052 CCGAAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1111
Oy 923 GCGTGTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 982
Db 1112 ACATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1171
Oy 983 GCGCGAGCGTGTGATGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015
Db 1172 CCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1204

```

## RESULT 13

AD39044 standard; DNA: 1248 BP.

AD39044:

23-SEP-2002 (first entry)

Streptomyces venezuelae Desi gene.

Glycosylated polyketide: modified recombinant bacterial host cell;

mRBHC; macrolide; anthracycline; angucycline; avermectin; milbemycin;

tetracycline; polyene; polyether; ansamycin; isochromanone; sugar;

desosamine; Desi; 4-dehydrase; enzyme; gene; ds.

Streptomyces venezuelae.

```

FH Key Location/Qualifiers
FT CDS 1..1248
FT /tag- a
FT /product- "Streptomyces venezuelae Desi protein"
FT /trans- except- (pos:1..3, aa:met)
FT /note- "CDS does not include start codon"
FT /partial

PN WO200229035-A2.
PD 11-APR-2002.
XX 05-OCT-2001; 2001WO-US31255.
XX 05-OCT-2000; 2000US-238185P.
XX (MINU) UNIV MINNESOTA.
XX (LIUH) LIU H.
XX (SHER) SHERMAN D H.
XX (ZHAO) ZHAO L.
XX Liu H, Sherman DH, Zhao L;
XX WPI: 2002-405171/43.
XX P-PSDB: AAE24229.
XX Modified recombinant bacterial host cells in which the expression and
XX activity of nucleic acids encoding sugar biosynthetic enzymes has been
XX altered, useful for producing metabolites with altered sugar structures
XX
XX Disclosure: Page 153-154; 174pp; English.
XX
XX The invention provides a method to alter the sugar structure diversity
XX for a particular metabolite via the recruitment and collaborative action
XX of sugar genes from a variety of sugar biosynthetic pathways to yield a
XX metabolite comprising a non-natural sugar, e.g., a novel glycosylated
XX polyketide. The invention also relates to a modified recombinant
XX bacterial host cell (mRBHC) in which the expression and activity of
XX nucleic acids encoding sugar biosynthetic enzymes has been altered.
XX The mRBHCs may be cultured to produce the modified sugar products.
XX e.g., a macrolide, anthracycline, angucycline, avermectin, milbemycin,
XX tetracycline, polyene, polyether, ansamycin or isochromanone.
XX The present sequence is Streptomyces venezuelae sugar (desosamine)
XX biosynthetic gene cluster Desi (4-dehydrase) gene.
XX
XX Sequence 1248 BP; 177 A; 546 C; 377 G; 148 T; 0 other;
XX
XX Query Match 4.7%; Score 81.8; DB 24; Length 1248;
XX Best Local Similarity 44.2%; Pred. No. 0.00015;
XX Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;
Oy 32 CCATGTCCCAATCCGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 91
Db 215 CCACCTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 274
Oy 92 ACGGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151
Db 275 AAGTATCATGCTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 334
Oy 152 CCATGCGGATCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 211
Db 335 TCACCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 394
Oy 212 TCGCGCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 271
Db 395 CCGCGCGCGGTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 454
Oy 272 ACATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 331
Db 455 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 514
Oy 332 GCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 391

```

(MINU ) UNIV MINNESOTA.

XX PI Sherman DH, Liu H, Xue Y, Zhao L;

XX WPI: 2000-160679/14.

DR P-PSDB; AAY71179.

XX

PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.

PT synthesis of methymycin and pikromycin -

XX

PS Claim 2; Page 281-287; 438pp; English.

CC The invention relates to an isolated and purified nucleic acid segment comprising a desosamine biosynthetic gene cluster, a fragment or its biologically active variant, where the nucleic acid sequence is not derived from the eryC gene cluster of *Saccharopolyspora erythraea* or Streptomyces antibioticus. The invention also relates to a macrolide biosynthetic gene cluster, or fragment thereof. The macrolide biosynthetic gene cluster encodes proteins which synthesise methymycin, pikromycin, neomethymycin, narboxmycin or a combination of these compounds. Recombinant or augmented cells comprising the desosamine and/or macrolide biosynthetic gene clusters are useful for the production of biologically active macrolides. The macrolide biosynthetic proteins are useful for synthesis of methymycin, pikromycin, neomethymycin and narboxmycin. The alternative termination of polypeptide synthesis may be useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA) monomers. The compounds produced by the recombinant host cells are useful as biopolymers, e.g., in packaging or biomedical applications, to engineer PHA monomer syntheses or to prepare biologically active agents, such as chemotherapeutics, immunosuppressants, agents to treat asthma, chronic obstructive pulmonary disease as well as other diseases involving respiratory inflammation, cholesterol-lowering agents or macrolide-based antibiotics which are active against a variety of organisms, e.g., bacteria, including multi-drug resistant pneumococci and other respiratory pathogens, as well as viral parasitic pathogens, or as crop protection agents (e.g., fungicides or insecticides) via expression of polypeptides in plants. The present sequence represents the desosamine biosynthetic gene cluster from *Streptomyces venezuelae* ATCC 15439.

Sequence 12441 BP: 1704 A; 4294 C; 4686 G; 1757 T; 0 other;

Query Match 4.7%; Score 81.8; DB 21; Length 12441;  
Best Local Similarity 44.2%; Pred. No. 0.00011;  
Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2.

32 CCATGTGCCAAATCGATTCCAGACCTTCGACTGCCTGGCGGCGCACCTCATCG 91  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 11982 CCACCTGGCAAGCCAGCGCGCGGGCTGCAGCTCTCGCGACGCGCGGCTCCACCGGG 11923

92 ACGGCAGCAACACC CGGGGGGGCGCGCTGCAGCTGGGGCTGGCGGCGAACCCGATGGCG 151  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 11922 AAGTATATCATGCGTGCGTAGTAAGATGTTGGCGCGCCACCCCGACGCACTGGCGCTGGATCGGGCG 11863

152 CCATGGGGCATGTGTCGCGAGCGCGCGCGCGCGCACCGCGGGTCGACGTCGTCGTCGGTGG 211  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 11862 TCACCCCGGCTCTTCGCGCGACATCGACCGCGAGAACCGGCAACCTCGAACCGGACAGAGTGG 11803

212 TCGCGCGCGGGCTTCATCGATCGATCGACACCCAGCAGACGAACAATACTGTCTCAAGCGCTGGC 271  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 11802 CGCGCGCGGTACACACCCCGCACCTCGGGCGCTGTGGCGGTCCACACTCTGTGGCGCGCGCCCT 11743

272 ACATGACGCGCCGAAGATTCGCGAGGGCGTCACACAGGTGTGTACAGGCGCAATTTCGGCATCA 331  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 11742 GCGCGCGCGACAGCTGCGGGAAGTGGCGGACGAGCGGCTGGCGCTGTACTTTCGAGG 11683

332 GCCTGGCGCGCTGGCGGACGCAACCCGCGCGCCCGCTGGACCTGTGAGACCTGTGAGAGCGG 391  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 11682 CGCGGACGCGCTCGGGCTCGGGGTCGAGGCGCGCGCGCGCGGACACTCGGCGAGCGCG 11623

392 GCTTCTACCGTTTCGAGCGCTTCGCGGACTACTGTGAGACCGCTTTGGGGGCGGACGCGGGG 451  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 11622 AGGCTTCAGCTTCCA--CGCACCAAGGCGGTCAACCGCTTTCGAGGCGGCGCGCGCTGG 11566





CC respiratory pathogens, as well as viral parasitic pathogens, or as crop  
CC protection agents (e.g., fungicides or insecticides) via expression of  
CC polyketides in plants. The present sequence represents the desmosmine  
CC biosynthetic gene cluster (pik) from *Streptomyces venezuelae* ATCC  
CC 15439, as given in figure 32.

XX  
SQ Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 other:

Query Match 4.7%; Score 81.8; DB 21; Length 13613;  
Best Local Similarity 44.2%; Pred. No. 0.00011;  
Matches 439; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

QY 32 CCATGTCCTCCATTCGATTCCTCCAGCCCTTCGACCTGCTCGCGGGGCGACCCCTCATCG 91  
DB 12669 CCACCTGCAACGCGACGCGGGCTCCAGCTCTCGCGGACGCGCGGCTTCACCGCGG 12610  
QY 92 ACGGCAAGCAACACCCCGGGGCGCGCCCACTGGGGCTGCGCGGACCGCATCGCG 151  
DB 12609 AAGTGATCATGCGGTGATGACCTTGCGCCGACCCCGCACGCACTGCGCTGATCGGCC 12550  
QY 152 CCATCGGCGATCTGTGAGCGCGCGGACACCCCGGCTGAGGTGTCGGGCTGTGG 211  
DB 12549 TCACCGCGGCTCTTGCGCGACATGACCCGGACACCGGCAACCTGACCCGACAGGTGG 12490  
QY 212 TCGCGCCCGGCTTCATGACTGACACCCAGACGACAACTACCTGCTCAGGCGTCGCG 271  
DB 12489 CCGCGCGGTGACACCCCGACCTCGGCGCTGCGGCTGCACTCGGCGCGCGCT 12430  
QY 272 ACATGACGCGCCAGATCTGCGAGGGGCTGACACGCTGCTCAGGGCATTTGCGGCATCA 331  
DB 12429 GCGCGCGCGACGAGCTGCGGAGAGTCCGACGACAGCGCTGCGGCTTACTTTCAGC 12370  
QY 332 GCGTGGCGCGCTGCGGCGACACCCCGCGCGCGCGCTGAGCCCTGCTGAGCAAGGCG 391  
DB 12369 CCGCGGACGCGCTCGGCTGCGGCTGCGAGCGCGCGCGCGCGCTGCGGCGACGCG 12310  
QY 392 GCTTTACGCTTTCGAGCGCTTGCAGCTGACGCTTGCAGCGGCGCAACGCGCGCG 451  
DB 12309 AGGCTTTACGCTTTCGAGCGCTTGCAGCTGACGCTTGCAGCGGCGCGCGCTGCG 12253  
QY 452 CCGTCAACGCGCGCTGATGATGCTGCGCATTCACGCTGCGCGCGCGCTGATGCGGACT 511  
DB 12252 TCACCGAGCGCGGCGCTGATGATGCTGCGCATTCACGCTGCGCGCGCGCTGATGCG 12193  
QY 512 TGCAGGCGCGCGCGCGCGACGAGGAAATGCGGCGCATGCGGACCTGCGGAGGAGCA 571  
DB 12192 TGCAGGCA 12133  
QY 572 TGGCGACGCA 631  
DB 12132 TGGCA 12073  
QY 632 CCACCGAAGATCATGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 691  
DB 12072 CCGCGTACCG 12013  
QY 692 CCCACATGCG 751  
DB 12012 ACGGCGTCAACG 11953  
QY 752 GCGCGAGCTGAGAGTCCGCGGTGATCTGCGACCAAGTCAAGTCAAGTCAAGTCAAGT 811  
DB 11952 ACGCGAGCTGAGAGTCCGCGGTGATCTGCGACCAAGTCAAGTCAAGTCAAGTCAAG 11893  
QY 812 TGGCG 862  
DB 11892 CCGCGGCGTCCGCGAGGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 11833  
QY 863 AGGAGCTCTGCTGAGCGCTATCTCCTACGCTGCGCGCGCTCACCATCTCAAGGAGGAC 922  
DB 11832 CCGAGGCTCTGCGCGCGCGCGGTGCTCCCTGCGCGAGCGCGCGCGCGCGCGCGCG 11773  
QY 923 GCGGCTGCTGCG 982

DB 11772 ACATCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11713  
QY 983 GCGCGGCG 1015  
DB 11712 CCG 11680

Search completed: January 14, 2003, 05:58:19  
Job time: 1874 secs



BASE COUNT 8270 a 25171 c 22273 g 9426 t  
ORIGIN

Query Match 5.6% Score 99.2; DB 6; Length 65140;  
Best Local Similarity 43.5% Pred. No. 0.0082;  
Matches 724; Conservative 0; Mismatches 913; Indels 27; Gaps 5;

QY 62 ACCTGCTGCTGCGGCGGACCCCTCATGAGCGGAGCAACCCGGGGGGCGGCG 121  
DB 19698 ACGACGGCTGCGGCGGACCCCTCATGAGCGGAGCAACCCGGGGGGCGGCG 19757  
QY 122 ACCTGCGGCTGCGGCGGACCCCTCATGAGCGGAGCAACCCGGGGGGCGGCG 181  
DB 19758 CCGGCGGCGGCGGACCCCTCATGAGCGGAGCAACCCGGGGGGCGGCG 19817  
QY 182 ACACCGGCGGCGGCTGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCG 241  
DB 19818 ACCCGCGGCGGCTGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 19877  
QY 242 ACGACGACACTGCTGCTGAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCG 301  
DB 19878 ACGCGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 19937  
QY 302 CCGCGGCTGCTGAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 361  
DB 19938 TCGCGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 19997  
QY 362 CCGCGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 421  
DB 19998 TCGCGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 20057  
QY 422 ACCTGAGCGCTTGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 481  
DB 20058 CCGCGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 20117  
QY 482 CAAGCGCTGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 541  
DB 20118 TCGAGCTGAGTGAACCCCGCTTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 20177  
QY 542 CGGCGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 601  
DB 20178 TCGTGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20237  
QY 602 GCGCGCTTCTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 661  
DB 20238 CCGCGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20297  
QY 662 CGCTGAGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 721  
DB 20298 TCGACCGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20357  
QY 722 TGGCGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 781  
DB 20358 CCGCGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20417  
QY 782 CGACGACAGCGTATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 841  
DB 20418 TGGTCTGCTGTAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20477  
QY 842 TCGAGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 901  
DB 20478 GCGTGAATCGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20537  
QY 902 CCACCATGCTCAAGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 961  
DB 20538 CCGACGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20591  
QY 962 AGCCCTTCCCGAACTGAGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1021  
DB 20592 ACGACGAGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20651  
QY 1022 CCAAGTGAAGCTGTGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1081

DB 20652 TCCGCTGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20711  
QY 1082 AACCGAGCTGACAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1141  
DB 20712 GCACCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20768  
QY 1142 CGACGAGAGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1201  
DB 20769 GCCACTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20828  
QY 1202 ATGCGCGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1261  
DB 20829 CCGAAGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20888  
QY 1262 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1321  
DB 20889 CCGG 20939  
QY 1322 TGGTGTGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1381  
DB 20940 TCGTGTGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20999  
QY 1382 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1441  
DB 21000 GCGTGAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21059  
QY 1442 TCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1501  
DB 21060 GCGACCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21119  
QY 1502 CCTTAC-----AATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1555  
DB 21120 CCGGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21179  
QY 1556 CAACCGCTACATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1615  
DB 21180 CCGCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21239  
QY 1616 AGAAGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1672  
DB 21240 TGACAGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21299  
QY 1673 AAGACTGTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1716  
DB 21300 TGACCTTGAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21343

Search completed: January 14, 2003, 08:07:30  
Job time : 5425 secs